

PCTWELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales BüroINTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation ⁶ : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027		A2	(11) Internationale Veröffentlichungsnummer: WO 99/33998 (43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)
<p>(21) Internationales Aktenzeichen: PCT/EP98/08216</p> <p>(22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98)</p> <p>(30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE</p> <p>(71) Anmelder (<i>für alle Bestimmungsstaaten ausser US</i>): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).</p> <p>(72) Erfinder; und</p> <p>(75) Erfinder/Anmelder (<i>nur für US</i>): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggendorfstrasse 59, D-51061 Köln (DE).</p> <p>(74) Gemeinsamer Vertreter: BAYER AKTIENGESELLSCHAFT; D-51368 Leverkusen (DE).</p>		<p>(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Veröffentlicht <i>Ohne internationalem Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>	
<p>(54) Titel: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF</p> <p>(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG</p> <p>A... FULL LENGTH</p>			
<p>(57) Abstract</p> <p>The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.</p>			

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Versinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	MN	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MG	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen Telomerase-Untereinheit und deren diagnostische und therapeutische Verwendung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten 10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der 30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

30 Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen 5 auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens 15 (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et* 20 *al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

25 Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt 30 werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgeneze ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein.
10 Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.
15

20 Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweligen Strahlen schädigen nicht nur die Tumorzellen, sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle
25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden
30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapierung von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-transkriptionale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein“) spezifisch gebunden werden.

20

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements) be-

30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

5 Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer 10 Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15 Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20 Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

25 Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

30 Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon 5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere 10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).

20 Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

30 Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, 5 M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise: 10

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle 15 Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

20 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

25 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

30 Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch 5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch 10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen. 15

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.

20 Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem. 25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilesequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

30 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

10 B. Messung des Substanzeffektes auf die Reportergenexpression.

15 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

20 Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

25 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

30 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,

10 B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;

15 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

20 Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Spraque Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des 10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

15 Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

20 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: 25 Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur 14: nicht verdaut.

30 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promoters vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

20

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

25

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

30

Exons sind als durchnummelierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymeschnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

10

15

20

25

30

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgesplicte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

15 Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite sind die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

20

25

Tab. 1: Exon Intron Übergänge des hTC-Gens

30 Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in
10 Intron 2.

Tab. 1

3' Acceptor Sequence		5' Donor Sequence									
Intron	Exon	Exon	bp	No.	Exon	bp	No.	Intron	Intron	Intr	bp
5' flankierende Region	cagggcgctcccccgtag	GTTTCAAGGAGGCTGCGT	1	281	CGCCCCCTCCTCCGCCAG	gtgggcctccccggggctcg	1	104			
	catgtcccttcgttttag	GTGTCCCTGCCCTGAGGAGC	2	1354	TGGCTGCCAGGACCCAG	gtgaggagggtggggccgt	2	8616			
	gagggggctctatttag	GGGTTGGCTGTGTCCGGC	3	196	TGCAAGGAGATGGAAATCAG	gtactgtatccccacgcca	3	2089			
	cccatgtctgtcccccgtag	ACAGCACTGAAAGGGGTG	4	181	GTTCCGAGAGAAAAGGG	gtggctgtcttttgttta	4	687			
	ctcgccatccactcacatcg	GCCGAGGGTCTACCTCGA	5	180	TGAGCTGTACTTTGTCAAG	gtgggtgcgggggaccccc	5	494			
	ccctctccatccatcgccatcg	GTGGATGTGACGGGCGGT	6	156	CAAGGCCCTCAAGGCCAC	gtaaaggttcacggtgtata	6	>4660			
	ctccatgtttttttttttcgtag	GTCTCTACCTTGACAGACC	7	96	TGCCGTGCTCATGGAGAG	gtctggcactgtgtgtca	7	98u			
	AGCTCTCTCCTGATGAGG	8	86	CCGTGCSCTCATGGGGCAA	gtgagtcagggtggccagggt	8	248:				
	GTCCTACGTCCAGGCCAG	9	114	CGGGGATTTGGGGACGG	gtggggccctttttccccc	9	1984				
	GCTGCTCCCTGCTTGGTG	10	72	ACGGGAAACCTTCTCTCAG	gtggggccgtgtgtgtgtg	10	1871				
	GACCTCTGGTCCGAGGTGTC	11	189	TGGAGGGGACTACTCCAG	gtgagcgccacctggccgga	11	380:				
	CTATGCCCGGCCCTCCATC	12	127	CCTGTTCTGGATTGGCAG	gtgaggcagggtgtatgtca	12	88n				
	GTGAACAGGCTTCAAGACGG	13	62	TCTCTGCTGAGGGGTACAG	gtgagccggccaccaagggg	13	318:				
	GTTTACCGCATGTGTGCTG	14	125	CTGAAAGGCCAAGACGAG	gtatgtgcagggtgtctggc	14	781				
	GGATGTCGGCTGGGGCCAA	15	138	CTGGGGTCACTCAGGACAG	gcaagggtgtgggtggggcc	15	536				
	CCCAAGCGCAAGCTGAGTC	16	664	TTTTTCAGTTTGTAAAAAA	3' flankierende Region						

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Spl	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
AP1	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der 5 Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

10

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen 15 bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer 20 Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, 25 Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschritte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

30

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 μ m; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 μ g/ml 10
15 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte 20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 25 als potentiell positiv heraus. Eine λ -DNA Präparation dieses Phagens Ausubel *et al.* (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen 30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) 5 gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach 10 λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umkloniert (Beispiel 4).

Beispiel 3

15 Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).
20 Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

25 Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.
30

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-
5 Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5'-Sequenz.

10 Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-
15 Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations- Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

20 Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

25 Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenkrons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenkron P17 und 5' von Phagenkron P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10

Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-

15

ACTATAAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

20

kloniert.

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATACC-3') zusammen mit der HDL-
5 Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers CSK-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen
15 genomicischen hTC-Region identifiziert (s. Fig. 7).

Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelnsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

5	ACTTGAGCCC	AAGAGTTCAA	GGCTACGGTG	AGCCATGATT	GCAACACCC	ACGCCAGCCT	TGGTGACAGA	70
	ATGAGACCT	GTCTCAAAA	AAAAAAA	AATTGAATA	ATATAAACGA	TCTTCTCTGG	CCACAGTGG	140
	ACAAAACCCAG	AAATCAACAA	CAAGAGGAAT	TTTGAAA	ACTACAAAC	ATGAAAATT	AACAATATAC	210
	TTCTGAATGA	CCAGTGAGTC	AATGAAGAAA	TTAAA	AAAGGAA	AATTGAAAAA	TTTATTAAAG	280
	CGGAAACATA	ACCTCTAAA	ACCCACGGTA	TACAGCAA	GCAGTGTCAA	GAAGGGAGTT	TATAGCTATA	350
10	AGCAGCTACA	TCAAAAAGT	AGAAAAGCCA	GCGCAGTGG	CTCATGCCTG	TAATCCCAGC	ACTTTGGGAG	420
	GCGAAAGCGG	CGAGATCGCC	TGAGGTCAAG	AGTTCGAGAC	CAGCCTGACC	AACACAGAGA	AACCTTGTCC	490
	CTACTAAAAA	TACAAAATT	GCTGGGCATG	GTGGCACATG	CTGTAACTCC	CAGCTACTCG	GGAGGCTGAG	560
	CGAGGATAAC	CGCTTGAAAC	CAGGGAGTGG	AGGTTGCGGT	GAGCCGGGAT	TGCGCCATTG	GACTCCAGCC	630
	TGGGTAACAA	GAGTGAACCT	CTGTCGAACT	AAAAAAA	AAGTGAACAT	ACTTTAAAAT	AACACCTAAT	700
	GATGCACTT	AAAAGACTAG	AAAAACCAAGA	GCRAA	ACTTAA	CTTAAAATTG	GTAAAGAAA	770
	AAGATCAGAG	CAGAAAATAA	TGAAAAGTAA	AGATAACAA	ACAAAAGATC	AACAAAATT	AAAGTTGGTT	840
	TTTGAAAAG	ATAAAACAAA	TTGACAAACC	TTGCCCAGA	CTAAGAAA	AGGAAGAAG	ACCTAAATAA	910
15	ATAAAGTCAG	AGATGAAAAA	AGACAGATTA	CAACTGATAC	CACAGAAA	CAAAGGATCA	CTAGAGGCTA	980
	CTATGAGCAA	CTGTACACTA	ATAAATTGAA	AAACCTAGAA	AAAATAGATA	AATTCTAGA	TGCAATACAAC	1050
	CTACCAAGAT	TGAACTGAA	AGAAATCCAA	AGCCCCAAAC	GACCAAAAC	AATAATGGGA	TTAAAGCCAT	1120
	AATAAAAAGT	CTCTTAGCAA	AGAGAACCCC	AGGACCAAAT	GCGCTCTCTG	CTGGATTTTA	CCAATCATT	1190
20	AAAGAACGAT	GAATTC	CCTCTAA	CTATTCTGAA	AAATAGAGGA	AGAAACTACT	CCAAACTCTAT	1260
	TCTACATGGC	CAGTATTACC	CTGATTC	ACCCAGACAA	AAACACATCA	AAACAAAACA	AACAAAAAAA	1330
	CAGAAAGAAA	GAAA	ACTACA	GGCAATATC	CTGTGAT	ACTGATACAA	AAATCTCAA	1400
	CCAAACCAA	TTAAACAA	CTCTGAAAG	ATCATT	CTGTGAT	GGGTTTATT	CCAGGGATGG	1470
25	AAGGATGGT	CAACATG	AAATCAATCA	ATGTGATACA	TCATCCCAC	AAATGAACT	ACAAAACAT	1540
	TATGATTATT	TCATTTATG	CAGAAAAGC	ATTGATAAA	ATTCTGCACC	CTTCATGATA	AAAACCTCA	1610
	AAAACCCAGG	TATACAAGA	ACATACAGCC	CAGGACAC	GGCTCAC	TGCGATCCC	GCAC	1680
	AGGCAAGGT	GGGATGATG	CTTGGGCCCC	GGAGTTGAG	ACTAGCCTGG	GCACAA	GAGACCTGTT	1750
	CTACAAAAAA	CTTTTTAA	AAATTAGCCA	GGCATGATG	CTATGCTG	TAGTCCCAGC	TAGTCTGGAG	1820
30	GCTGAGGTG	GAGAATCACT	TAAGGCTTAGG	AGGTCGAGG	TGCACTGAGC	CATGAACT	TCACTGTACT	1890
	CCAGCTAGA	CCAAAGAAC	AGACCCACT	GAATAAAGA	AAGGAGAAGG	AGAAGGGAGA	AGGGAGGGAG	1960
	AGGGAGGAG	GAGGAGAAGG	AGGAGGTTG	GGAGAAGTGG	AGAGGAAAGG	GGAGGGAAAG	GAGGAAGAAG	2030
	AAAGAACATA	TTTCACACATA	ATAAAAGGCC	TATATGACAG	ACCGAGGAG	TATTATGAGG	AAAAGCTGAA	2100
	AGCCCTTCC	CTAACATG	GGAAATGACA	AGGGCCACT	TTCACACTG	TGATTCACAA	TAGTACTAGA	2170
35	AGTCCTAGCT	AGAGCAATCA	GATAAGAGAA	AGAAATAAA	GGCATCCAA	CTGGAAAGGA	AGAAGTCAAA	2240
	TTATCCTGTT	TGCAAGATG	ATGATCTT	ATCTGAA	GACTTAAAG	ACCATTA	AACATTAGA	2310
	GCTGAATT	GGTACAGCAG	GATAAAAT	CAATG	AAATCACTG	TATTCTATA	TTCCACACG	2380
	AAACATCTG	AAAAGAAC	CAAAAAGCA	GCTACAAATA	AAATTAAACA	GCTAGGAAT	AACCAAAGAA	2450
40	GTGAAAGATC	TCTACAATGA	AAACTATAA	ATGTTGATAA	AAAGAAATTG	AGAGGCCAC	AAAAGGAA	2520
	AGATATTCCA	TGTTCATAGA	TTGGAAGAAT	AAATACTGTT	AAATGTC	TACTACCAAA	AGCAATTAC	2590
	AAATTCAATG	CAATCCCTAT	AAAAACTA	ATGACGTTCT	TCACAGAA	AGAGAAACA	ATTCTAAGAT	2660
	TTGTACAGAA	CCACAAAAGA	CCCAGAATG	CCAAAGCTAT	CTCTGACCA	AAGAACAAA	CTGGAAGCAT	2730
	CACATTACCT	GACTTCAAT	TATACAA	AGCTATAGA	ACCCAAACTA	CATGTTACTG	GCAT	2800
	AGATGAGACA	TGGACCAAGAG	GAACAGAATA	GAGAATCCAG	AAACAAATCC	ATGCATCTAC	ACTGAACTCA	2870
45	TTTTGACAA	AGGTGCAAG	ACACATCTT	GGGGAAAAG	TAATCTCTT	AATAATGTT	GCTGGAGGAA	2940
	CTGGATATCC	ATATG	AAACATACT	AGAAGCT	CTCTCCACAT	ATACAAAAGC	AAATCAAAAT	3010
	GGATGAAAGG	CTTAAATCTA	AAACCTCAA	CTTGC	ACTAAAGAA	ACACGGGAG	AAACTCTCCA	3080
	GGACATTGGA	GTGGGCAAG	ACTTCTGAG	TAATTCCCTG	CAGGACAC	CAACAAAGC	AAAACAGAC	3150
	AAATGGGATC	ATATCAAGT	AAAAGCTTC	CCCGACG	AGGAAACAAAT	CAACAAAGAG	AGAGACACAA	3220
50	CCACAGAAATG	GGGAAATATA	TTTGC	AACTCATA	TAAGGAA	ACCTTGGGAC	ACTGTTGGTG	3290
	GCTCAACTA	CTCTATAAGA	AAAACCTCA	ATAGCTGAT	TTTC	AAAGGAA	TCTGGGATAGA	3360
	CATTCTCAA	AAATAAGTCAT	AAACATGCA	AAACGGCATC	TGAAAATGTC	CTCACACCA	CTGATCATCA	3430
	GAGAAATGCA	AATCAAAACT	ACTATGAGAG	ATCATCTCAT	CCCAGTTAA	ATGGCTTTTA	TTCAAAAGAC	3500
	AGGCAATAAAC	AAATGCGAT	GGGGCAAG	GAGGATGTTG	TTAAAGGAA	ACCTTGGGAC	ACTGTTGGTG	3570
55	TTGCTACCAC	TATGGAGAC	AGTTGAAAG	TTCTC	AAACTA	AAAGCTTACCA	TACAGCAAT	3640
	CCATTGCTAG	GTATATAC	CAAAAAGGG	AATCACTGTA	TCAACAACT	ATCTCCACTC	CCACATTAC	3710
	TGCAGCACTG	TTCATAGC	CCAAAGTTG	GAAGCACCT	CAGTGTCCAT	CAACAGACGA	ATGGAAAAAG	3780
	AAAATGTTGG	GCACATCAC	ATGGAGTAC	TACGCGACCA	TAAAAGAA	TGAGATCTTG	TCAGTGTCAA	3850
	CAGCATGGGG	GGCAACTGTTG	AGTATGTTAA	GTGAAATAAG	CCAGGACAC	AAGAACAAAC	TTTCATGTT	3920
60	CTCCCTTACT	TGTTGGGACCA	AAAAATAAA	CAATTGACAT	CCCAGTTAA	ATGGCTTTTA	TTCAAAAGAC	3990
	GGTGGGGGAC	AGGGT	GACTA	TAATTGACAT	AGAAATAGAG	GAGAATGGTG	TTCTAGAGG	4060
	TGTTGTTT	GTAAACACAA	GAAGGATAA	ATGCTGAG	GTGACAGATA	CCCTGATTTAC	CCTGATGTGA	4130
	TTATTACACA	TTGTATGCT	ATGAT	AAACATTA	TTGCTATAGAT	AAACCCCTA	CTATATTAA	4200
	AAATTAATT	TTAATGGCCA	GGCACGGT	CTCATGTC	TAATCCCAGC	ACTTTGGGAG	GGCGAGGCGG	4270
65	GTGGATCACC	TGAGGTCAGG	AGTTGAAAG	CCAGTGGCC	ACCATGATGAT	AACTCTGCT	CTACTAAAGA	4340
	TACAAAATT	AGCCAGGCGT	GGGTTTCTA	TTAGACGCTC	TAATGTTAAAG	AACTCTGCT	AAATTAAACC	4410
	TTGCTTGAAC	CTGGGAGGG	GAGGTGCG	TGAGCCGAGA	ACCTGTAGTC	CCAACTACTC	AGGAGGCTGA	4480
	GAGCAAGACT	CCATCTCAA	ACAAAACAA	AAAAAAGAG	ATTTAAATTG	TAATTTTAT	GTACCCGATA	4550
	AAATATAC	CTACTAT	AGAAGTTAA	AAATTAAACA	ATTTAAAG	GTAATTAACC	ACTTAATCTA	4620
70	AAATAAGAAC	AATGTATG	GGGTTTCTA	TTCTCTGAAGA	AGTAAAAGT	ATGGCCACGA	TGGCAGAAAT	4690
	GTGAGGAGG	AAACATG	GGGTTTCTA	TTAGACGCTC	TAATCTCTG	TAATGTTAC	AAATTAAACC	4760
	AAAGACAGGC	TGGGAGGAGT	TAAGAGGCA	TCTCTATAAGC	CCTAAACAA	CTGCTAATAA	TGGTGAAGG	4830
	TAATCTCTAT	TAATTACAA	TAATTACAGA	TATCTCTAA	ATCGAGCTG	AGAATGGGA	CGTCTGATCA	4900
	CACCGTCCTC	TCATTCAGG	GGCTTTTTT	CTTGTGTTGCT	TGGAGATT	CGATTGTTG	TTCTGTTGTT	4970
	CTTAAACTA	ATCTGTATGA	ATCC	AAACAA	AAAAATGGT	GGTGTATTCC	TCCAGAAGAA	5040
75	TGGCAGGAAG	CAGGTGGC	TGTGGACCTG	AGCCACTCA	ATCTTCAGG	GTCTCTGGCC	AAGACCCAGG	5110

5	TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT 5180
	CGTTGGTAG CAGCGCATGA AGTGCCTTA TTACGCTT GCAAAGATTG CTCTGGATAC CATCTGGAAA 5250
	AGGGGGCAG CGGGAAATGCA AGGAGTCAGA AGCCTCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA 5320
	CCCGGGCGTG TGCCAGAGG AGAGGACTCA AGGCACCTCG AGATATGGCT TAATCTTTT TTTCACCTGAA 5390
	AGCAGTGACC AAGGTGTTATT CTGAGGGAGG CTTAGAGTTAG GTGCCCTCTT TAAACAGAA AGTCATGGA 5460
	GCACCCCTCT GCAGGGAAAA CCAGACGCC GCTCTGCGGT CATTACCTC TTTCCTCTCT CCCTCTCTTG 5530
	CCCTCGGGT TTCTGATCGG GACAGACTGA CCCCGCTGGA GCTCTCCCGA GCGCGTGCCTG AGGACCCCTCT 5600
10	TGCAAGGGC CTCACAGACC CCCGGCTGG AGAGAGGAGT CTGAGGCTGG CTTAATAACA AACTGGATG 5670
	TGGCTGGGG CGGACAGCGA CGGGGGATT CAAAGACTTA ATTCCATGAG TAAATCAAC CTTTCCACAT 5740
	CGGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAATTT TCATCAAATA ACATTCAGGA CTGCAGAAAT 5810
	CCAAGGGC AAAACAGGA CTGAGCTATG TTGGCAAGG TCCAAGGACT TAAATAACCAT GTTCAGAGGG 5880
	ATTTTTCGCC CTAAGTACTT TTATTTGGTT TCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGGAGG 5950
	AGAGGGCTGG CGGGCAGGGC TATGAGCAGC GCGAGGCCA CGGGAGAGA GTCCCCCGGCC TGGGAGGCTG 6020
15	ACAGCAGGAC CACTGACCGT CCTCCCTGG AGCTGCCACA TGGGGCAACAG CACCCCTGGT 6090
	GTGACTCAGG ACCCCATACC GGCTTCTGG GCCCCCACAG ACTAACCCAG GAAGTCACGG AGCTCTGARC 6160
	CCGTGGAAAC GAACATGACC CTGGCTGCC TGCTCTCTG GGTGGCTCAA GGTTAATGAA GTGGTGTGCA 6230
	GGAATGGCC ATGTAATTAA CAGACTCTG CTGATGGGG ECGGCTCTTC CATCATTATT CATTTACCC 6300
	CCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAGGCAATGA CAAAACTCAG TACAAACACC 6370
20	ACTCTTTAC TAGGGCCACA GAGCAGCGG CACACCCCTG ATATACATTAG AGTCAGGAG AGATGAGGCT 6440
	GCTTCAGCC ACCAGGCTGG GGTGACAACA CGGCGTGAAC AGTCTGTTCTC TTAGACTAG TAGACCCCTGG 6510
	CAGGGACTCC CCCAGATTCT AGGGCCCTGG TGCTGCTTCC CGAGGGCCGC ACTGCCCCCTG GAGACTCAGG 6580
	CTGGGTGCCC ACACTGAGGC CAGCCCTGTC TCCACACCCCT CGCCCTCCAG GCTCCAGCTT CTCCAGCAGC 6650
	TTCTTAAACC CTGGGCTGGC CGTGTTCAGG CGCTACTGTC TACCTGTC TTGTCACCGG 6720
25	ACCTAGTCG CACGGTCTCT CTCACATGG GGTGCTGTGTC TCTTCTCCCA ACACTCACAT GCGTTGAAGG 6790
	GAGGAGATTC TGCGCCCTCC AGACTGGCTC CTCTGAGGCC GAACTCGGTG CTTGGATGGT GATGCAGGTT 6860
	CCTGGCGTCC CGGTCGACGC TGACCTCCAT TTCCAGGGC CCCCCCTGTC TTGTCATCTG CGGGGGCTG 6930
	CCGGTGTGTT CTCTGTTTC TGTCGCTCTT TCCACGTCAGA GCTCTGTTG TCTCTGCCCG CTAGGGCTCTC 7000
	GGGGTTTTA TAGGCACTAGG CAGGGGGCTG GGTGGCTCAA GGAATGCAA CATTGGGTG 7070
30	TGAAAGTAGG AGTGCCTGTC CTACCTAGG TCAACGGGCA CAGGGCTGGG GATGGAGCCC GCGCCAGGG 7140
	CCCCCCTTC CTCGCCCAGC ACCTTCTCTG CCCCCCTCCCT CTGGAAACACA GAGTGGCAGT TTCCACAAAGC 7210
	ACTAAAGCAT CTCTCCCAA AAAGACCCAGC ATGGGACCC CTGGACATT GCCCCACAGC CCTGGGAATT 7280
	CACGTGACTA CGCACATCAT GTACACACTC CGGTCGACGAGA CGGGCCCGG CTGTTTATT TTAATAGCTA 7350
	CAAAGCGAGG AAATCCCTG TAAATACAAAC TGGTTAAACAA AGGGGTCCA TCCGCACGGT 7420
35	GGAGCAGTCC TCACAGTGA GAGGAACATG CGGTTTATAA AGCCTCAGG CATCTCAAGG GAATTACGGT 7490
	GAGTCAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTTC ACCCCATGCC 7560
	AGGGGAGTGG TTAGGGGGTTA TAAGGAGCTG GGGGGCGGCA GCTGGGGCTG ACTGCACGCA CCTTTTACTA 7630
	AAGCCAGTTC CCTGGTTCTG ATGGTATGG TCTAGTTATG GAGAGTAAC TATAGGGAG TGGGGATGGG 7700
40	GGAGCCCGGA GGCTGTGCCA TCTTGGCCAT GCCCCGAGTGT CCTGGGAGG ATAATGCTCT AGAGATGCC 7770
	ACGTCCTGAT TACCCCAAAC CTGTCGACAG ACCCCAGGGC CTTGGCAGGT GTGATCTCG 7840
	TGAGGACCTC GAGGCTCTGG ATCTCTGGG ACTACCTGCA GGGCCAAAAA GTATTCAGG GGTCTGGGA 7910
	AGAGGGCGGC AGGAGGGTCA GAGGGGGCA GCTCTCAGGAC GATGGAGGCA GTCACTGTC GGTCTGAAAG 7980
	GGAGGGAGGG CTCCTCGCAAG GAGCCCTGCAAG GGGGCTCCAG AAGCTGAAAG AAGCAGGGAA GGGACCCCTCC 8050
45	ACCGAGCTG CGCAGGAAAG CGACGGCTGG CCCCCCTGGT CTAGCATGAA GTGTTGGGG ATTGCGAA GCAACAGGAA 8120
	GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTTGGGG ATTGCGAA GCAACAGGAA 8190
	ACCCATGCACT GTGAACTCTA GGATTATTC AAAACAAAGG TTACAGAAA ACATCCAGG CAGGGCTGAA 8260
	GTGCTCTCGG CGAAGGGCAG GCAGGGCAGC AGTGTATTA TTAGCTATT TTATTTATT TACTTACTTT 8330
	CTGAGACAGA TTATGCTCT TGTTGCCAG GCTGGAGTGC AGCGGCATGA TCTGGCTCA CTGCAACCTC 8400
50	CGTCTCTGG GTCAGCAAGA TTCTCGTGC TCACTGCAAGA ATCTGAGCTG GATTTCAGGC GTGCAACACC 8470
	ACACCCGCTC AATTGTTAT TTTAGTAGA GATGGGCTT CACCATGTTG GTCAAGCTG TCTCAAAATC 8540
	CTGACCTCG AGTGCATGCC CACCTCAGCC TCCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT 8610
	GGCCTATTAA AACATTCTG GGTCTCAAGT ACACCAACTG GTAAGGAGTT CATGGAGTTC 8680
	AATTCCCTT TTACTCAGGA GTTACCTCC TTTGATATT TCTGTAATT TTCTGAGACT GGGGATACAC 8750
55	CGTCTCTGTA CATATTCAAA GTTCTCTGTA CCACCTGTTA TCCCATGGGA CCCACTGAG GGGCAGCTGG 8820
	GAGGCTGAG CTCCTAGGTC CCACTGTTGTT AGCTCATCTG CAGTGAACAA CTAGTGTAGA ATCAGGGCC 8890
	AAGTGTGAC ATCTGCTGA ATCTCAATG CTCACTGTTG GTCAAGAACAT TTAGAAATTAA AGTCTCATCC 8960
	CTCTCTACTC ATCTGGATTG AGCCCCCTCC CTATCCCCC CCAGGGGAGG AGGAGTTCTC CTACTCTCTG 9030
60	TGGAGGAAGG ATGTGACTT TGTATTTT CACTCTGTTG ACTGAAATCA CTGTTTCAATT GTGTTGTTTG 9100
	TTTGTGTTGT TTGAGAGGC GGTTCACACT GGTGCTGCA CGCTGGAGGG AGTGCATGG CGGCATCTTG 9170
	GCTTACTGCA GCCTCTGCCG CCCAGGTCA AGTGTATCTC CTGCTCTCCG CTCCCATTTG GCTGGGATTA 9240
	CAGGCACCCG CACCATGCCG CAGCTAATT TTTGATTTT TCTCTCTCT TCTCTCTCT CTTTAAAT TGTGTTTCT 9310
	ATGTTGGCCA GCGCTGCTC GAACCTCTGA CCTCAGATGA TCTGGCTCTT CTTTAAAT TGTGTTTCT 9380
	GATTACAGGT GTGAGCCACC ATGGCCAGGT CAGAAATTAC TCTGTTTAAAC AACATCTGGG TCTGAGGTAG 9450
65	GAAGCTCACC CCACTCAAGT GTTGTGTTG TTTAGGCCA TGTAGAAATT TTTTATTTGT GTTGAACAA 9520
	CTCTTGATGTT TTACACTGT GTAGACTAAAG ACATCATCAG CTTTCAAAAG CACACAAAC TGACCCATCA 9590
	ATACTGGGGT TACCTGGG TATCAGCAAT CTTCATTGAA GGGGGGAGG CGTTTCTCTC CCATGCACAT 9660
	GGTGTAAATT ACTCCAGCAT AATCTCTG TCTCTCTCT TCTCTCTCT CTTTAAAT TGTGTTTCT 9730
	ATGTTGGCTT CTCTGCAAG ACCAGCTGTA AGCTACAAC TAACTTTGT TGAACAAAT TTCCAAACACC 9800
70	GCCCCCTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCCCTTAAAG AGGCTTAGGG ATCAACTAAAGG 9870
	GGATTTCTAG AAGAGGGACG TGTAATCTA AGTATTCAGA AGACGAGGCT AACCTCCAGC GAGCGTGAACA 9940
	GCCCCAGGGAG GGTGGAGGC CTGTTCAAT GCTAGCTCA TAAATAAGG AATTCTCTC GGCAGTTCT 10010
	GAAAGTAGGA AAGGTTACAT TTAAGGTTGG GTTGTGTTAGC ATTCAGTGT TTGCGACCT CAGCTACAGC 10080
	ATCCCTGCAA GGCGCTCGGA GACCCAGAG CTTCTCGCC CTTAGATCC AAACCTGAGC AACCCGGACT 10150
75	CTGGATTCTC GGGAAAGTCT CAGCTGCTCT GGGTGGAAAGT CGGGCTCTCT AGCTCTGCAC TCCGGAGCTT 10220
	CCGCTGTGGCT TCTACTGCTG GGGTGGAAAGT CGGGCTCTCT AGCTCTGCAC TCCGGAGCTT 10290
	CCGCTGGACCC CGAGGGCTGCC CTCCACCTG TGCGGGCGGG ATGTGACAGG ATGTTGGCTC CATCTGCCAG 10360
	ACAGAGTGC CGGGGCCAGG GTCAAGGCG TGTGTTGCTGG TCTGAGGCGC CGGTGCGCG GCGACGAGGA 10430
	GGCGCTGGCT CCATTTCTCA CGGGGACCC CGGGGGGGT GATTAACAGA TTGGGGTGG 10500

5	TTTGTCTATG	GTGGGGACCC	CTCGCCGCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCCTG	TGTCAAGGGAG	10570
	CCCAAGTCGC	GGGGAAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCCCCTCCAG	GGAGCAATGC	10640
	GTCCCTCGGGT	TCGTCCTCAG	CGCGCTCTAC	GCGCTCCGT	CCTCCCTTC	ACGTCCGGCA	TTCGTGGTC	10710
	CCGGAGCCCG	ACGGCCCGCG	TCCGGACCTG	GAGGCACTGC	TGGGTCCTCG	GATCAGGCCA	GCGGCCAAAG	10780
	GGTCGCCGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCCCTCCCT	GGGGTTACCC	CACAGCTTAG	10850
	GCCGATTCTGA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCTC	CCTGCACCC	GGGAGCGCGA	GCGGCCGCGG	10920
	GGCGGGGAAAG	CGCGGCCCG	CTGGGGCCCT	CGCTGGCTC	TGTGCGCAG	CCACTACCCG	GAGGTGCTGC	10990
	CCAGTGGATT	CGCGGGCACAC	GACGCCAGG	ACCGCCCTCC	CCACGTGGCG	GAGGGACTGG	GGACCCGGGC	11060
10	ACCCGTCTG	CCCCCTCACC	TTCCAGCTCC	GCCTCCCTCG	CGGGGACCC	GCCCCGTC	GACCCCTCCC	11130
	GGGTCCCCGG	CCCAGCCCC	TCCGGGCCCC	CCACGCCCC	TCCCTTCTCC	TCCGGGGCCC	CGCCCTCTCC	11200
	TCGGCGCCG	AGTTTCAGGC	AGCGCTCGCT	CTCTGTCGC	ACGTGGAAAG	CCCTGGCCCC	GGCCACCCCC	11270
	CGGATGCCG	CGCTCCCCG	CTGCCAGGCC	GTGCGCTCCC	TGTGCGCAG	CCACTACCCG	GAGGTGCTGC	11340
	CGCTGGGACAC	GTTCGTGCC	CGCGCTGGGG	CCACGGCTC	GGCGCTGGGG	CAGCGCGGG	ACCCGGCGGC	11410
15	TTTCCCGCCG	CTGGTGGCC	AGTGGCTGTT	GTGCGTGGCC	TGGGACTGGAC	GGGGCCCCCCC	GGCCGCCCCC	11480
	TCTCTCCGCG	AGGTTGGCC	CCCCGGGGCT	GGCGTCCGGC	TGGGTTGAG	GGCGCCGGGG	GGGAACCAAGC	11550
	GACATCGGGA	GAGCAGCGCA	GGCGACTCAG	GGCGCTCCC	CCCGAGGCTT	CCTGCTGAA	GGAGCTGGTG	11620
	CCCCGAGTGC	TGCGAGGCT	GTGCGAGGCC	GGCGCAGAAGA	ACGTGGCTGC	CTTCGGCTTC	GGCGTCTGG	11690
	ACGGGGGCCCC	CGGGGGCCCC	CCCGAGGCTC	TCACCAACAG	CGTGCAGCAG	TACCTGCCCA	ACACGGTGAC	11760
20	CGACGCACTG	GGGGGGAGGC	GGGGCTGGGG	SCTCTGCTG	CGCCGGCTG	GCGACGACGT	GCTGGTTTAC	11830
	CTGCTGGCAC	GCTGCGCTC	TTTGTGCTG	GTGGCTCCCA	GTGTCGCTTA	CCAGGTGTC	GGGGCCCCCG	11900
	TGTACCAAGCT	CGGGGCTGCC	ACTCAGGGCT	GGGGCCCGGC	ACACGGCTAGT	GGACCCCGRAA	GGCGCTCTGG	11970
	ATGCGAACCG	GCCCTGAACC	ATAGCGCTAG	GGAGGGCGG	GTCCTCTGG	GCTGCCAGC	CCCGGGTGGC	12040
	AGGAGGCGCG	GGGGCAGTGC	CAGCGGAAGT	CTGCGTTCG	CCAAAGGCGC	CAGGGCTGCC	GTCGCCCCCT	12110
25	TGGTTCTGT	GTGTTGTCAC	GTGCCAGACC	GGCGCAAGAA	CCACCTCTT	TGGAGGGTGC	GCTCTCTG	12250
	ACCGGCCACT	CCCACCCATC	CGTGGGGCC	CAGCACACC	CAGGGCCCCC	ATCCACATCG	GGGCCAACAC	12320
	GTCCCCTGGG	CACCGCTTGT	CCCCGGGTG	ACGCGAGAC	CAAGCCTACT	CTCTACTCTC	CAGGGGACAA	12390
	GGAGCAGCTG	CGGGCCCTCT	TCTACTCAG	CTCTCTGAGG	CCACGCTG	CTGGCGCTCG	GAGGCTCTG	12460
30	GAGACCATCT	TTCTGGGTT	CAGGCCCTGG	ATGCCAGGG	CTGGCCCGAG	FTGCCCCCCC	CTGCCCCAGC	12530
	GCTACTGGCC	AAATCGGGCC	CTGTTTCTGG	AGCTGCTTGC	GAACCAAGGC	CAGTGGCCCT	ACGGGGTGTCT	12600
	CCTCAAGACG	CACTCGCCG	TGCGAGCTGC	GGTCACCCCA	CAACCGGGT	TCTGTGCCCG	GGAGAAGCCC	12670
	CAGGGCTCTG	TGGCGGCC	CGAGGAGGAG	GACACAGACC	CCCCTGCGCT	GGTGAGCTG	CTCCGCCAGC	12740
35	ACAGCAGGCC	CTGGCAGCTG	TACGGCTCTG	TCGGGCC	CGAGCTGAGC	CTGGGCCCC	CAGGGCTCTG	12810
	GGGCTCAGG	CACAAACGAC	GGCGCTCTC	CAGGAACACC	AAGAAGTCA	TCTCTCTGG	GAAGCATGCC	12880
	AAGCTCTGC	TGCAAGGAGC	GACGTGGAAG	ATAGCGCTGC	GGGACTGCGC	TTGGCTGCG	AGGAGCCAG	12950
	GTGAGGAGGT	GGTGGCCCTC	GAGGGCCCG	GGCCCAAGAG	TGAATGCTG	AGGGGCTCAG	AAGAGGGGGC	13020
	AGGCAGACGC	CTGGCTCTCG	TGTCCTCATC	TCACCTGGG	CCACAGTGGC	TTTTCGCTCA	GGACGTGAG	13090
40	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGCTATA	ACTTACGAGG	TTCACCTTC	13160
	CGTTTTGATG	GACACGCGC	TTCCAGGGCT	CGAGGCGCAG	CGAGTGAAC	TACTTGGGAG	GCTGAGGCTG	13230
	TGAGGGGGGG	TGCGGCAA	TTGGGAGAAG	TGCTGAGG	CACAGACCT	CTGGGAGGG	TCGCTGAGG	13300
	TTACCTATAA	TCCTCTTCG	ARTTCAAGG	GTGGAATGA	GAGGTGGGG	CGAGAACCCC	CTCTCTCTGG	13370
	GGGTGGGGAG	TAAGGGTTT	GCAGGTGAC	GTGGTCAGG	AAATGTCAGG	TTTGTGTTTA	AGATTAAATT	13440
45	GTGTGTTGAC	GGCCAGGTGC	GGTGGCTCAC	GGCGGTAACT	CCACGACTT	GGGAAGCTGA	GGCAGTGG	13510
	TCACCTGAGG	TCAGGAGTT	GAGACCAAGCC	TGACCAACAT	GTTGAAACCC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGGTG	TGTGTCCTG	TAATCCCG	TACTTGGGAG	GCTGAGGGCTG	GAGGAATCACT	13650
	TGAACCCAGG	AGGGCGAGGC	TGAGATTG	TCATTGACT	CCAGCTGGG	CGACAAAGAT	13720	
50	AAAGACTCTG	CTTTAAAGAA	AAAAGTGT	CGTTGATTGT	GGCAGGAGAC	GSTAGAGGG	GGGAGATAAG	13790
	ACTGTTCTCC	AGCACAGATC	CTGGTCCCAT	CTTGGTAGGT	GAAGGGAGGC	ACATGGGAGC	AGGGACACG	13860
	AGATGGCTCC	ACCTGCTGAG	GAAGGGAGAC	TGTTTG	GGGGCTGGG	ATGGTGTCTC	TGGGCCCTGC	13930
	CGTGTCCCCC	CCCTGTTTT	CTGGATTGTA	TGTTGAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGCTCCAG	14000
	TGCTCCCGAG	CCCTACCGT	CGAGCTAGA	GAAGTCCCAG	TTTCACCCAA	CTCCAAAGAC	14070	
55	ATGTAAGACT	TCCGGCCATG	CGACAAAGGA	GGGTGACCTT	CTTGGGCCCC	TTTTTTCTT	TTTTTTCTT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTG	AGTGCAGAAT	14210
	TGCTAACTCG	GGCGGGTTTA	CAGCAGGTT	CTTGGAAATG	TGCGTC	GTGACTGGAA	GTCCCCTACCC	14280
	ATCGAAGCGC	AGTCCTCTCA	CACCTGCTGC	GGCTCAGGTG	GACCCAGCGC	AGTCAGATAA	GGCTCATGCA	14350
	ACCCAGTTT	GCTTTTTG	CTCCAGCTC	CTTCAGTGG	GAGAGTTGAG	TTTCTCTG	CAGGACTCTG	14420
	CTTGTCTATG	ACTTCAGATG	AGGTCAACAT	CTGGGCTG	CTTATGCAAG	CTGGGCTG	CTTATGCAAG	14490
	TGGTCCCCGG	GTGTCCTGT	CACGTGCAAGG	GTGAGTGGAGG	CTTGGGCCCC	AGGTGCTCC	GTACCTGTA	14560
60	GGGTGAGTGA	GGCGGGCTCC	CTGGGTGTC	CTGGGCTG	CAGGTGAT	GAGGTGTG	CCCCGGGTG	14630
	CCCTGTCAAG	TGAGGGTGA	GTGAGGGCC	ATCCCCGGT	GTCCCCTGCA	CCTGTTGAGGT	GAGTGGGG	14700
	TGGTCCCCGG	GTGTCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACTGTC	GGGTGTCCT	GTACCTGCA	14770
	GGGTGAGTGA	GGCGGGCTCC	CTGGGTGTC	CTCTCAGGT	TAGGGTGTG	GAGGGGAGGC	CCCCAGGGT	14840
	CCCTGTCAAG	TGAGGGTGA	GTGAGGGCC	GTCCCCTGGG	GTCCCCTCCCA	GTTATAGGGT	GAGTGGGCA	14910
65	CTGTCCCCGG	GTGTCCTGT	CACGTGCAAGG	GTGAGTGAGG	CGGGGCCCC	GGGTGCTCC	CTCAGCTGCA	14980
	GGGTGAGTGA	GGCGGGCTCC	CTGGGTGTC	CTGTCTCGT	TAGGGTGTG	GAGGGTCTGT	CCCCAGGGT	15050
	CTTGGGGTT	TGCTCACTG	AGCTTCTCC	TAATGTTG	CTCTTCTAT	GGCCACAGCT	GGCCGGTTG	15120
	CCCATGGCT	GGGTAGATGG	TGCAAGGC	GTGCTGTC	CCAAGCTAT	CTTTCTGAT	GTCGGCTCT	15190
	TCTTGGTCAC	CTCTCCGTC	CATTGTCGA	GGGGGACACG	GGACTGCAAG	CTCTGCCCTC	CCCGTGC	15260
	GGCACTGCA	CCACAGCTTC	AGGTCCGCT	GCCCTCTG	GGCTATAGGGT	GAGTGGGCA	15330	
70	ATCCATGCTG	CCAATACTCC	TCTCCAGCT	TGTCATGC	CGAGGCTGGA	CTCTGGCTG	CCTGTGCTG	15400
	CTGCCACGTG	TTGCTGGAGA	CATCCAGAGA	AGGGTCTCT	GTGCTGTC	GGAAAGCAAG	TCACCCCAAGC	15470
	CCCCCTCACTT	GTCCTGTTT	CTCCCAAGCT	GCCCCCTCTG	TTGGGCCCC	TGGGGTGGGT	GAACAGCTG	15540
	TCACCTTATT	CTGGGCACCT	GGCGCTCATT	GCTTAGGCTG	GGCTCTG	CCAGTGC	CCTCACATGG	15610
	ATTGACGTCC	AGCCACAGGT	TGGAGTGTCT	CTGTCATG	CCTGCTG	GACCCACGTG	GAGGGCCGGT	15680
75	GTCTCCGCA	GGCTCTGCA	GACTCTCTC	TTGGGCTTA	GTTTGAAT	TCACTGATT	ACCTCTGACG	15750
	TTTCTATCTC	TCCATTGAT	GCTTTTCTT	GGTTTATTCT	TTCATTCTT	TTCTAGCTTC	TTAGTTAGT	15820
	CATGCCCTTC	CCTCTAAGTG	CTGCGCTTAC	TGCAACCTG	TTTTGATG	GAAGTAATCT	CAACATCAGC	15890

5	CACTTTCAAG TGTCTTAAA ATACTTCAA CTGTTAACAT TTCTTTAAAG TATTCTTATT CTGTGATT 15960 TTCTTGTG CACGCTGTG TTGACGTGA AATCATTG ATATCAGTGA CTTTTAAGTA TTCTTACGT 16030 TATTCGTGA TTCTTGTGAG CAGTGAGTT TTGAAACACT TTATCAGTGA CTTTTAAGTA TTCTTACGT 16100 ATACGGTAGG TATTTTAAGT TATCATTAA TTATGATTG CTAAACTCAGT TGTTGAGTGG TCTGTATAAT 16170 ACCAATTATT TGAAAGTTGC GGAGCCTTGC TTGTTGATCT ATGTTGTCGA TTGTTTCCAG AACTGTCCAT 16240 TGAAATTAG ACATCCGTG AATAGTGGGC ATGATGTTG ACTATATCCA GCTTATAAG GTCCAGTGC 16310 AAGCTTGTG CTCTCTTAG ATGCAAGAAA TTCCAAGAAG GAGGCCATAG TCCCTCACCT GGGGGATGGG 16380 TCTGTCATT TCTCTCGTT TGGTAGCATT ATGTTGAGG ATGTTAGGT CTATGCCAGT GGTAAGATT 16450 TTATCTTCTG GATGAGTGA TCTTTGGAG ACTCTATGT CTCTAGTAAT CTAGTAATT TTTTTTAAA 16520 TTGGCTTCTAG TACTGCCACA CTGGCTTCTG TTGTTAGGT ATTTTCTCTG TTCTGCTTGT 16590 AATTATATA TATATATAA TTTTTTTT TTGAGACA GAGTCTGGT CTGTCGCCA GGGTGAGTGC 16660 AGTGGTGTGA TCACAGGTCA GTGTAACATT TACCTTCTGG CCTGAGCGT CCTCTCACCT CAGCCTCTG 16730 AGTAGCTGGA ACTGCAGACA CGCACCGCTA CACCTGGCTA ATTTTTAAAT TTGTTCTGGA GACAGGGTCT 16800 TGCTGTGTTG CCCAGGTGG TCTCAAACTC TTGGAACATCA GGGATCACATC TACCTCGCT TCCCAAGTG 16870 CTGAAATTACA GGCATGAGCC ACCATGCTG GCCTAAATTAA CAACACTTT ATATTCTTAT AGTGTGGGT 16940 TGTCTGTGA ACAGCATGTA GGTAAGTTT CAATCCAGTC ATGCTGCTGTG TTCTGCTTGT 17010 TTTATTCTCA TTTTTTGTG ACTAGAGACC CGCCCTGGT ACCTCTGTTT TCCACTTGC TTGTCATGT 17080 CCTCGTCTCC TTGTTCTCA CCACCTCTG GTTGGCCATG TCCGTTTCTT GCGGAGTGTG TTGTCATCT 17150 CTCGTGTGCTT CTGGTCACT GGGCATTTGG TTGATCTTTT AAAATGAAGT CTGAAACATT GCTACCTTG 17570 20 ATTGTCGTG TTGCTTTG TTTATGAGA CAGTCCTACT CTGTCACCA GGGTGGAGTG TAATGCCACA 17220 ATCTCGGCTC ACTGCAACCT CTGCTCCCTC GGTCRAAGCA GTTCTCATTC CTCAACCTCA TGAGTAGCTG 17360 GGATTACAGG CGCCCAACCC CACCGCTGGC TAATTTTTGT ATTTTGTAGG GATAGTGGCTT TTACCATG 17430 TGGCCAGGCT GGTCTAACAC TCTGACCTCA AGTGTCTG CCGCCCTGGT CCGCCACAG TGCTGGATT 17500 ACAGGTGCAA GCCACCCCTGC CGGGCATACCC TTGATCTTTT AAAATGAAGT CTGAAACATT GCTACCTTG 17640 TCCCTGAGCA TAAGACCCCT AGTGTATTTT AGCTCTGGG TGTTACCCCTC CTGTCGTGCTG TTTCCTCTG 17720 TGACTTAGT CTATCTCAGG CATCTTGACA CCCCCAACAG CTAAAGCTTA TAAATATTGT TTCCGTGTT 17710 GAGTGTGTTG GTAGCTTGC CCCCCCTG CTTTCTCC TTTGTTCCC GTCTGCTTC TGTCAGGGC 17780 CCGCGCTGCTG GGGTCCCCCT CTTGTCCTC TGCGTGGTTC TTCTGCTTC TTATGCTG 17850 30 CTTTACCTGT GCTGGCTTCC ATGGCATCTA CGGAGCTGGC GGGACCTCTG CTATGATGC ACAGATGA 17920 ATGTGGAGAC TCACGAGGAG GGCGGTCTAC TTGGCCCTG AGTGTCTGGA GCACCCACCTG GCCAGGTT 17990 CTTAGCCAGT GAGTGCACGC AACGCTCCGT CGGGCTGGG TGACCTGGA AAACCCAGG CATGTCGGGG 18060 TCTGGTGTGCCC CCCGGTGTG GAGTTGAAA TGCCCAACAG CTGCGTGTG CGCCGAGCTC TGACGGTGT 18130 GCCCTGGCGG GGAAGTGTG CTTCTCCCT TCTGTTGGG AACCAAGGACA AAGGATGAGG CTCCGAGCCG 18200 TTGTCGCCCA ACAGGAGCAT GACGTGAGGC ATGTTGATTA TTTTTAAAT TCTAGGCTGG CGCGCGTGGC 18270 35 TCACCCCTGT ATACCCAGCA CTTTGGAGG CCAAGCCGGG TGATCACGAG GGTAGGAGG TGAGACCAT 18340 CCTGGCCAAC ATGATGAAAC CCCATCTGT ACAAACAC AAAAATTAGC TGGGGCTGGT GGCGGGTGCC 18410 TGAAATCCA GCTACTGGG AGGCTGAGGC AGGAGAATTG CTGACCTGGA GGAGTTGAA TGTCAGTGA 18480 CCCGACATG CCAACTCTGC TTCCAGCTG CAAACACAGC GAGACTCTGT CTCAAAAAAA AAAAAAAA 18550 40 AAAAAAAA AATTCTAGT GCCACATTA AAAAGTAAA AAAGAAAGGT GAAATTAAATG TAATAATAGA 18620 TTTACTGAA GCCCAGCATC TCCACACCTC ATCATTTAG GGTGTTATTG GTGGGAGCAT CACTCRACGG 18690 ACATTGACA TTTTTGTGCG GGATCTGGT TGAGGCTGGC TGCGTGGCC ATCTGGCT 18760 GGACCTGCTG GGCTTCCCAT GGCATGGGTT GTGTTACCG AGTGTGCAAGG TCCGGGATGAA GTTCGCCAGG 18830 CCCTCAGTGA CTTGGATGTG CAGTGTCTGG ATGGTGCACC TGTCGGATG TGTCGGATG 18900 45 AGCTGGATGT GTGGTGTG TGAGTGTGAG CTTGGGGTGC AGGTCCTCAG GCCCTCGGT AGCTGGAGGT 18970 ATGGAGTCCC GTAGTGTGAG GTCGGGGTGC AGGTCCTCAG GCCCTCGGT GAGCTGGATG TGTCGGTGT 19040 GGATGGTGA CGTCAGGGGT GAGGTCTCA GGCCCTCGGT AGGCTGGAGG TATGGAGTCC GGATGATGCA 19110 GGTCGGGGGT GAGGTGCCA GGCCCTGTG TGAGTGGAT GTGTTGTC TGATGGTGC AGGTCTGGGG 19180 TGAGGTCACC AGGCCCTGGC TGAGCTGGG TGAGCTGGT CAGGGTGGAA TGAGGTCGCC 19250 50 CAGACGGTGC CAGACCATGC TGAGCTGG ATATGGGGT CTCGGATGGT CGAGCTGGT GGTCAGGTCTG 19320 CCAGGCCCTG CTGTGAGTTG GATGTGGGT GTCCGGATGC TGAGCTGGC GTGTGAGGT ACCAGGCCCT 19390 CTGTGAGGT GGATGTGTTG TGTCGGATG TGTCAGGGT GGGCTGGG TGAGCTGGC CTCGGTGTG 19460 AGCTGGATGT GTGGTGTG TGAGTGTGAG GTCGGAGTC AGGTCGGAG CCGCCCTGGT AGCTGGATGT 19530 GCAGTGTCCA GATGGTGCAG GTCCGGGGTG AGGTCGGCAG ACCCTCGGT GAGCTGGATG TGCGGTGTCT 19600 55 GGATGGTGA GGCTGGTGGAT GAGGTGCCA GGCCCTGGT GAGCTGGATG TATGGAGTCC GGATGTTGCC 19670 GGTCTGGGGT GAGGTGCCA GACCCCTGTG TGAGCTGGAT TGTCGGTGTG TGATGGTAC AGGTCTGGAG 19740 TGAGGTCGCC AGACCCCTGT TGAGCTGGA TATGGGGT CCGGATGGT CAGGTCAAGG TGAGGTCCTC 19810 CAGGGCTCC TGAGCTGGG GTGATGGATG CGGATGATG TGTCGGGG TGAGGTGCG CAGGCCCTGC 19880 TGTGAATGGG ATGTCGGGG TGTCGGATGGT CGAGGTCTGG TGTCGGGG TGAGGTGAGG TGTCAGGTCTC 19950 60 AGGTATGGAG TCCGGATGAT GCAGGTCGGG GGTGAGGTGG CCAGGCCCTG CTGTGAGCTG GATGTGCGGC 20020 GTCTGGATGG TGCAAGTGTG GGGTGTGGG GCGACCCCTG CGGTGAGGT GAGGTATGGA TCCGGATG 20090 TGCAAGTCCC GGGTGGAGGT GCGACCCCTG TGTCGGATGG CAGGTGGCTG TATCGGATG TGTCAGTCCC 20160 GGGTGAGGTG GCGACCCCTG CTGTGAGGT GGTATGTCG TATCGGATG TGTCAGGTCT GGGGTGAGGT 20230 CACCAGGCC CTCGGTGGC TGTTGTCGG TGTCGGGGT GTCAGGGTCA CGGGGGTGAAGT TGCCAGGCC 20300 65 CTCGGTGGC TGGATGTGCG TGTCGGGGT GTCAGGTCTG GGGTGGAGGT GCGACCCCTG TGTCAGGTCTG 20370 TGTTGGGCTG GATCTGGCTG TGTCGGATGG TGCAAGTGTG GGGTGGAGGT GCGACCCCTG TGTCAGGTCTG 20440 GATGTGCGGT GTCTGATGG TGCAAGTGTG GGGTGGAGGT GCGACCCCTG TGTCGGGGCTG GATGTGCGGT 20510 GTCTGGATGG TGCAAGTGTG GGGTGGAGGT GCGACCCCTG TGTCGGATGG GAGATGTCGG TGTCAGGTCTG 20580 70 GTGCAGGTCC GGGGTGAGGT AGCCAAGGCC TTGGGGTGGAG TGAGATGTCG TGGGGTGGG GTGTCGGGGAT GTGCAAGGTCTC 20650 CGGGGTGAGG TGCCGAGGCC TGCGCAGGCC TGCGGGTTAG CTGGATATGC TGTCGGGG TGTCAGGTCTG 20720 GTCACCAAGGC CCTGGGGTAA GCTGGATGTC CGGTGTCGG ATGGTGCAGG TCCGGGGTGA GTGCGCCAGG 20790 CCCTGCTGTG AGCTGGATGT GCTGTATCCC GATGGTGCAG GTCCGGGGTG AGGTGGCCAG GCGCTGCAGT 20860 GAGCTGGATG TGCTGTATCC GGTGAGGTCA GGTCTGGGT GAGGTGCCA GCGCCCTGGG TTAGCTGGAT TGTCAGTCCC 20930 ATGCGGTGTC GGATGGTCA GGTCGGGGTG GAGGTCAACCA GCGCCCTGGG TTAGCTGGAT GTTCAGTCCC 21000 75 CGGGATGGTGC AGGTCTGGGG TGAGGTGGC AGGCCCCCTGT GAGGTGGAGG TGAGCTGGG TGTCAGGTCTG 21070 CAGGTGCCGG TGAGGTGCG CAGGCCCTG CGGTGAGCTG GATGTGAGT GTACGGATGG TGCAAGTCCC GGGTGAGGT 21140 CGTGAGGTGCG CAGGCCCTG CGGTGAGCTG GATGTGAGT GTACGGATGG TGCAAGTCCC GGGTGAGGT 21210 GCCAGGCCCT CGGGTGGGCT GTATGTCGTG TGTCGGATG GTGCAAGTCCC GGGTGAGGT CGGGAGCC 21280
---	--

5	TGCGGTGAGC	TGGATGTGTG	GTGCTGGAT	GTCAGGTC	CGGGGTGAGT	TGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCC	GTGTCCCCGT	GTCCGAATGG	TGCAAGGTC	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATGTGCCGT	GTCCGGATGG	TGCAAGGTC	GGGTGAGGTC	GCCAGGCCCT	TGGTGAGCTG	GATGTGCCGT	21490
	GTCCGGATGG	TGCAAGGTC	GGGTGAGGTC	ACCCGGCCCT	CGGTGATCTG	GATGTGCCAT	GTCTCTCTCG	21560
	TTAAAGGGGT	TGGCTGTGTT	CGGGCCGAG	AGCACCGCT	CGGTGAGGAG	ATCCTGGCA	AGTCTCTGCA	21630
	CTGGCTGATG	AGTGTGTAGC	TGCTGGAGT	GTCAGGTC	TTCTTTATG	TCACGGAGAC	CACGTTCAA	21700
	AAGAACAGC	TCTTTCTA	CGGAAAGT	TGCTGGAGCA	AGTGTCAAG	CATTGGAAATC	AGGTACTGTG	21770
	TCCCAACGCC	AGGCCCTCTC	TTCTCGAAGT	CGTGGAAAC	CAGCCCGCC	TCAGCATGCC	CCTGTCTCCA	21840
	CTTGGCTGTG	CTTCCCTGCC	TGTCAGGTC	TGGGCTGGGA	GCCAGGGGCC	CGTCACAGG	CCTGGTCCAA	21910
10	GTGGATGCC	TGCAAGGTC	TGACTGCTG	GAGCTCACGT	TCTCTTACTT	GTAAATCAG	GAGTTTGTG	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAACGAGA	GGGATTTAAA	TTAGATGAA	ACACTACAC	TAGCCTCTT	22050
	GCCTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTTTT	TTTTTTTTT	TTGAGATGG	AGTCTCACT	22120
	TGTTGGCCAA	TGGGAGTGC	AGTGGCATAA	TCTGGCTCA	CTGAAACCTC	CACCTCTGG	GTTTAACGCA	22190
	TTCACAGCC	TCAGCTCTC	AAGTACGTG	GATTACAGGC	ACCTGCAC	ACGCCCTGGT	AATTGTTGTA	22260
15	CTTTTAGGAG	AGACGGGGTT	TCACCATGTA	GGCCAGGCTG	GTCTCGAATC	CATGACCTCA	GGTGATCCAC	22330
	CCACCTGGC	CTCCCCAAGT	GCTGGGTTA	CAGGCTAACG	CACCGTGC	AGCCCCGAT	TCTCTTTAA	22400
	TTCATGCTG	TCTGTATGAA	TCTCTAATCT	ATGGGATTTA	GGTCATGAGA	GGATAAAAATC	CCACCCACTT	22470
	GGCGACTCAC	TGCAAGGAGC	ACCTGTCAG	GGAGCACCTG	GGGATAGGG	AGTCTCCACCA	TGAGTAACT	22540
20	TCTAGGGTGC	TGGCTGTGAA	TGGCTGTG	ATTTGCTG	CAATGTC	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTC	AAGCTGGATT	TGCACTAGT	AGGGACGGGA	GGCGTGTCT	GGGAGATGCC	AGCCTGGCT	22680
	AGCCCCAGCC	ATGGTATTAG	CTTCTCGT	GAGCTGGGG	GCTGACTGTG	GAGGGCTTTA	GTCAAGAAGAT	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACTCG	AGTCCCCTGG	GGGCTCTG	ACACCCATG	CCCCAAATCA	22820
	GGATGCTGC	AGAGGGAGCT	GGCAGCAGAC	CTCGTCAGAG	GTAAACAGC	CTCTGGCTG	GGGACCCCGA	22890
25	CTGTTGGCTC	GGGGCATTTC	CTTGCATCTG	GGGGAGGGTC	AGGGGCTCT	CTGTTGGAAAC	AAGTTAATAC	22960
	ACATGCAAC	TACTTAGAC	TTTACACGTA	TTTATGTTG	TTGACCCAA	CATGGTCATT	TGACCACTAT	23030
	TTTGGAAAGA	ATTTAATTGG	GGTACCGGA	AGGAGCAGAC	AGACGTGGT	GTCCCCAAGA	TGCTCTTGT	23100
	CACTACTGG	ACTGTTAGTC	GGCTGGGGG	GGCTGGGAGG	CCCTCTCTC	CTGGACAGGG	TACCGTGCCT	23170
30	TTTCTACTCT	GCTGGGCTCG	CGGGCTCGG	TCAAGGACCC	AGTCTCGAG	CACCCCGGC	CCCCAGTCTC	23240
	ACGGAGTGC	AGGCTGTCA	CCACAGATGC	CCAGGTCAG	GTGTTGGC	TCCAGCCCC	GTGCCCCCAT	23310
	GGGTGGTTT	GGGGGAAAAG	GGCAAGGCA	GGAGGTCTCG	GAGACTGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCTCTGGC	TGAGCTGCC	TGAGACGCT	CTCCCCCCT	CTCCATCTG	AGGGATGTGG	CTCTTTCTAC	23450
	CTGGGGTTC	TGCCTGGGG	CAGCCTTGGG	CTACCCCTAGT	GGCTGTACCA	GAGGGACAGG	CATCTGTGT	23520
	GGAGGGCCT	GGGGTCACTG	GGCCAGGAGT	CGACCTGGGG	ACCAGCTCTC	CTGGTGTGTA	TGTTGGGACA	23590
35	GTCAACCTGG	GGGGTGGAGC	CGGGACTGGG	TGGCTGGGG	GTTGACTATA	GGACCAAGTG	TCCAGGTGCC	23660
	CTGCAACTAG	AGGGGCTCTC	AGAGGCGTCT	GGCTGGCATG	GGTGGACGTG	GCCCCGGCA	TGGCCTTCAG	23730
	CGTGTGCTGC	CGTGGGGTGC	CTGAGTCCTG	ACTGAGTCGG	TGGGGCTTG	TGGCTTCCC	TGAGCTTCCC	23800
	CCTAGTGTG	TGTCCTGGT	AGCAAGCCTC	CTGAGGGGGT	CTCTATTGCA	GACAGCACTT	GAAGAGGGTG	23870
	CAGCTGCGG	AGCTGTCGGA	AGCAGAGGTC	AGGCACTAC	GGGAAGGCCAG	GCCCCTGCTG	CTGACGTCCA	23940
40	GACTCCGAC	CATCCCCAAG	CTTGCATGCA	GGAGCTGGGG	GATGACATG	GACTACGTG	TGTTGGGCAAG	24010
	AACTCCCGC	AGAGAAAAAGA	GGGGTCTGT	GCTTTGGTTT	AACCTCTT	TTAAACAGAA	GTGCTTGTG	24080
	GCCCCACATT	TGGTATCAGC	TTAGATGAAG	GGCCCGGGAGG	GGGGGCACG	GGACACAGC	AGGGCCATGG	24150
	CACGGCGCCA	ACCCATTGTTG	GGCGCACAGT	AGGTGGCCG	GGTGGCGGTG	CCTCCAGAAA	AGCAGCGTGG	24220
	GGGTGTAGGG	GGAGCTCTGT	GGGCAAGGAC	AGGCTCTGAG	GACACCAAGA	AGCAGCCGG	CCAGGGCTG	24290
45	GATGCACAC	GGCCCGAGGT	CTTGGATCCG	TGTCCTGCTG	TGGTGC	GAGGCGCAG	CCTCCGTGCG	24360
	CGGGGGCG	GGACAGGCC	ACGACTGCCA	GGAGCCACCC	GGGGCTCTG	GATCTGGGAG	TGTTGGGAC	24430
	GGCTCTCTGCA	GGGGCACCCCC	TGGCTGCGG	TGGCTGGGT	GACCCCTGCA	TCTGAGGAGA	GTGTTGGGTG	24500
	AGGTGTGAG	AGGTGTGAGCA	TGAGGATCCC	TGTTGCAACA	CACATGCC	CAGGAACCCG	TTTCAACAG	24570
	GGTCTGAGGA	AGCTGGGAGG	GGTCTAGGT	CCGGGGCTG	GGTGTG	GACACTGGGG	AGGGGCTGCT	24640
50	TCTCTCTGG	GTCTCTATGG	TGGGGTGGG	ACTGGGGCGG	ACTCACTTC	CTGACTGTCT	CCCCATGCTG	24710
	CCCCGGCAGG	CCGAGCTGCT	CACCTCTGAGG	TGTAAGGCAC	TGTTGAGGGT	GCTCAACTAC	GAGGGGGCG	24780
	GGGGGGGG	CTCTCTGGG	GGCTCTGTC	TGGGGCTGGA	CGATATCAC	AGGGGCTGGG	GCACCTCTG	24850
	GCTGCTGTG	CGGGCCGAGG	ACCCGGCC	TGAGCTGTAC	TTTGTCAAGG	TGGGGCCGG	GGACCCCCGT	24920
55	GAGCCACCT	CTGTCACCTT	GGGAGTGGCT	GGCTGATTGG	CACCTCTAGT	TGGGGTGGAGG	AGGTAACCTCT	24990
	GGGTGGGG	GGGGAGTGGC	AGGTGACCT	GGTCACTG	AGGACACACC	TGGCACCTAG	GTTGGAGGCC	25060
	TTCAGCTTT	CCTGCAGCAC	ATGGTACCTG	CTGTCACCC	TGACTGCGG	GGCTCTTATT	CCAAGGGAGG	25130
	GTCCCACACTG	ATTCCTAGTT	CGCTCAGAGA	AGGAACCCGA	ACGGCTCTAC	CACCAAGCCC	CGGTGCTT	25200
	CACCCAGTC	CTGACCCAGG	GGTCTCTGT	CTGGGGGAC	AGAGGGGA	CACAGCCCG	CCTGCCCC	25270
	GGGTCTGGAG	TGGTGGGGGT	CGAGAGAGA	GTGGGGGACA	CGGGCAGGCC	AGGCCCTGAC	GGCAGAGGTG	25340
60	ATGTCTGAGT	TTCTGCTG	CCACTCTGTC	TCTCTCTGCG	TCTCACTACA	CAGGTGGATG	TGACGGGGCG	25410
	GTACGACACC	ATCCCCCAGG	ACAGGCTCAC	GGAGGTCTAC	GGCAGCATC	TAACACCCA	GAACACCTAC	25480
	TGCGTGCCTC	GGTATGCCGT	GGTCCAGAAG	GGCCGGCCATG	GGCACGCTCG	CAAGGCTTC	AAAGGCCACG	25550
	TAAGGTTCA	CTGTCATAGT	CGTGTCCAGG	ATGTGTCT	CTGGGATATTG	ATGTTGCTCA	GAATGCACT	25620
	GTGCTGTG	TGCGTTCTG	TGGTGGAGGT	ACTTCCATG	TTACACATG	TGTGATATG	GTGTTGCGA	25690
65	CGTGTGTG	GTGGTGCATG	TATCTGTC	GTGCAATT	GTGGTGTG	TGTGTTG	ACGTGTGTT	25760
	CCATGGTGTG	TGTGCTGTG	GTGTCATG	GTGTTG	GTGACACCTG	CATGTTCATG	CTGTTGCTG	25830
	CATGCTCTG	ATGTCCTAT	TGTCGGTGT	TGTCATG	ATATGGCTG	CTATGGCATG	25900	
	GGTCTGTG	GGCCCTTGGC	CTTACTCTT	CCTCTCTCAG	GCATGGTCCG	CACCATGTC	CTCACGCTCT	25970
	CGGGTGTG	TTTGGGGAGC	TCCACATCA	GGGTCCTCAG	TTCTAGCATG	GGTGCCTCTG	TCCGTGACA	26040
70	GGGGTGGGCC	TTGGAGACTG	TAGCCAGGT	TGAGAGGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCCTGGAC	26110
	CCCTGGCACC	CCCAGGACCC	CACTCTGCC	TATGCCGCT	CCATGAGATA	TAGGAAGGCT	GATTCAAGCC	26180
	TCGCTCCCCG	GGACACACTC	CTTCCAGAGC	GGGGGGGGG	CTTGGGGCTC	GGCAGGGGGT	AAAGGGGCC	26250
	TGGGCTTGGG	TTCCCAACCA	GTGGTCA	GGGGTAAGCC	CTAAAGTCG	TGCAAGGCC	TGCAAGGCC	26320
	GGGTGCAAGG	GTGAAGAAGT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCA	CATGTGAAA	CCACAAAGGA	26390
	CCTTTCTC	TGACTCTT	AGCT					26414

Contig 2:

5 TGTGGGATTG GTTTCATGT GTGGGATAGG TGGGGATCTG TGGGATTGGT TTTTATGAGT GGGGTAACAC 70
 AGAGTTCAAG GCGAGCTTC TTCTGTAGT GGGTCTGCAG GTGCTCCAC AGCTTTATTG AGGAGACCAT 140
 ATCTCCCTT GAACTATGGT CGGGTTTATA GTAAAGTCAGG GGTGTTGGAGG CCTCCCTGG GCTCCCTGTT 210
 CTGTTCTTC CACTCTGGG TCGTGTGGT CCTGCTGTGG TGTTGTCAGG CCGTCCTGG AATTCCCTG CGAGTTGGAG 280
 TCCCTTGTGT CATTGGCTG GATGTGGCC TGGCTACGT CGCTCCTGG TGATTAACAGA GTCTCGCTCT TTTTGGCCA 350
 GCTTCTTTC TTCTTTTTT TTTTTTTTT TGATTAACAGA GTCTCGCTCT TTTTGGCCA 420
 GGCTGGAGTG GTTGGGCTG ATCTTGGCTC ACTGCAACT GTGCTTCTG AGTTCAAGCA ATTCTCTTGC 490
 10 CTCAGCCCTC CAAAGTACGTG GAATTTATAGG CGGCCACAC CATGCTGACT AATTTTGTGA ATTATAGTAG 560
 AGACGAGGT GCTCCATGT GGCGGAGCTG GTCTGAACT CCTGACCTCA GTGATCCTC CCACCTCGGC 630
 CTCCAAAAGT GCTGGGATGA CAGTGTGAA CGCCGGCGGC CGCCGGAGAC TCGCTTCTG CAGCTTCCG 700
 GAGATCTGCA CGGATAGCTC CCTGAGCTGAC AACCTCCGTT TTCCCTCTCC AGGTCTCGCT 770
 15 AGGGGTCTT CATTCTATG ACTCTCTTCA CAGAAGAGTT TGACCTGTG TGATTTCCCG GCTGTTTCTC 840
 GCGTAATTGG TGTCTGTGT TTATCGATGG CCTCCCTTCA TTCCCTTCTG GCTTGTGTTA TTGTTGTTT 910
 TCCGGCTCTT CGAAGAAA GTTGGATTA TGGATGTTTG ATATATCTTCT TCTAAACAAG CATCTGAAGT 980
 TGCCGGTTTC CCTCTAAGG AGGATACCCG AGGGCCCTG CTGTTGGAGT GCACCCGCTC GGGGCTCTT 1050
 AGGAACCCGG CGCACAGCGG GAGGCTAGGT GGGGTGTGGG GAGCCAGCGT TCCCGCTGAA GCCCCGCCCC 1120
 20 TCTCAGATCA CGAGTGGCAT CGCGGATGCA GAGGGCACA CACCCCTACTG AGAACTGTG TGAGAGGGGG 1190
 TCTAGATCTT GTGCTCTTA TGGGAATCTA ATGCCATGATG ATCTGAGGTG GAAACGTTTG CTCCCAAAC 1260
 CATCCCCCTC CCCACTGCTG TCTCTGTGAA AAATCTGCTT CCTGCTCTG GACGAAACCC AGTCCCTGGT 1330
 TTGGGGACCC TGTGCTAAAG ACCTGCTTCA CGACGGCTCTC GTCTGCTGTT ATATATTTGGC TTTCTGTGT 1400
 TGAGTCCAGA ATAAATTACGG ATCTCTGTGA TCTCTTCCG CGACCTCAGA CCCATGGGCT ATTGTGGGC 1470
 GTGTGGCTG CTCTGGTTT GGGAAAGGGT CAGGCCCAT GTACCTCTT GTTACTGCTC TCCAGTTGG 1540
 TTCTCAGGGT TGAATCTGAC TCTGATGTTG TTAGGCCCA CGCCCTGGCC CGAGCTCTG GGGGCTGGGG 1610
 AACATGCTGA ACCACAGAGT CACCGTGGC TCTCTTGTAT GCTCTCAGA CTCGAGGCCT CCTGCTGCGG 1680
 TGTTAGTGTG TGTACGTGCG CGCTGTGAA CCTGCTCTGG GGACGCGAGG GCTTAGCAGG TCCCGTAGTA 1750
 AATGACAAAGC GTCTGGGGGG AGTCTGCGAGA ATAGGAGGTG GGGGCTGGCC TCTCTCTTCA CGCTCTTCA 1820
 25 ACTCTCTTCC TGCCTGTGCT GTGGCTGAC GCAATCCCTC CAGCACTGGG CTGGAGAGGC 1890
 CCGGGAGCTC GRGTGCCACT TGTGCCACGT GACTCTGGAT GGCAGTCGGT CAGGGGGTC TGATGTGTG 1960
 TGACTGTGGA TTGGCTGGTCA TCAGGGGGT CGCTGATGTC TGACTGTG TGAGCTCTG TGGGGCTCTGA 2030
 TGTGGTACT GTGGATGGCG GTCTGGGGGT CTGATGTG TGACTGTG TGAGCTGATG GACGGCTGCG 2100
 TGTGGTACT GTGGATGGCG GTCTGGGGGT CTGATGTG TGACTGTG TGAGCTGATG GGGGCTGATG 2170
 30 TGGTGTACT GTGGATGGCGT GTGGGGTCT GATGTGTG TGACTGTG TGAGCTGATG GGGGCTGCTG 2240
 CGTGGACTGT GTGGGGTCT GATGTGTG TGACTGTG TGAGCTGATG GGGGCTGCTG GGGGCTGATG 2310
 TGTGGTACT GTGGATGGCG GTCTGGGGT CTGATGTG TGACTGTG TGAGCTGATG GGGGCTGCTG 2380
 TGTTGGTGTG CTGTTGGATGG CGGCTGGGGT CTGATGTG TGACTGTG TGAGCTGATG GGGGCTGCTG 2450
 TGTGGTGTG TGATGGTGTG CAGGGGTCTG ATGTGTGGT TGACTGTGATG GGGGCTGCTG 2520
 35 GGTCTGATGT GTGGTACTGT TGGATGGTGA TCGGTACAG GGGTCTGATG TGTTGACT GTGGATGGG 2590
 GTCTGATGTG TGACTGTG TGATGGTGTG ATGGGGGGT TGCTGATGTTG TGCTGATGTTG TGACTGTG 2660
 GATGGCAGTC GTCGGGGTCT GTGGTACTGT GATGTGTG TGAGCTGGG CTGGGGGCT GATGTGTGTT 2730
 GACTGTGGAT GCGGGCTGTG GGGTCTGATG TGTGGTACT GTGGATGGG TGCTGGGGT CTGATGTGTT 2800
 GACTGTGGAT GCGGGCTGTG GGGTCTGATG TGTGGTACT GTGGATGGG TGCTGGGGT CTGATGTGTT 2870
 40 GACTGTGGAT GCGGGGGTGTG ATGTGTGGT TGACTGTG TGAGCTGATG GGGGCTGATG 2940
 GTGGTACTGT TGGATGGCAG TCGTGGGGTC TGATGTGTG TGACTGTG TGAGCTGATG GGGGCTGATG 3010
 GTCTGGTGTG TGTGGATGCC GGTCTGGGGT TCTGATGTG TGACTGTG TGAGCTGATG GGGGCTGCTG 3080
 ATGTGTGTG ACTGTGGATG CGCTCTGGG GTCTGATGTG TGAGCTGATG GGGGCTGCTG 3150
 ATGTGTGTG ACTGTGGATG GTGATCGTC ACAGGGGGT GATGTGTG TGACTGTGATG GGGGCTGCTG 3220
 45 GGGTCTGTG TGCTGGTACT GTGGATGGC GGTCTGGGGT CTGATGTG TGACTGTG TGAGCTGATG GGGGCTGATG 3290
 GGGTCTGTG TGCTGGTACT GTGGATGGC GGTCTGGGGT CTGATGTG TGACTGTG TGAGCTGATG GGGGCTGATG 3360
 GTCAACAGGG TCTGATGTG GTGACTGTG GATGGGGCTG GTGGGGCTG ATGTGTGGTG ACTGTGGATG 3430
 GCGGGCTGTG GGTCTGTG TGCTGGTACTG TGATGTG TGCTGGGGTC TGATGTG TGACTGTGGA 3500
 TGGGGCTGTG GGGGTCTGAT GGTCTGACTG TGATGTG TGCTGGGGTC TGCTGACAC GGGTCTGATG TGTTGTTGACT 3570
 50 GCAGGGTGGAG TCCCAAGTGTG ACTTTGCTGCTC CTGGGCCCCCT CGGGCCCCCGT TTCCAAAACAA 3640
 GAAGCTTCCC AGGGCGCTCTC TGGCTTCTAT CGGGCCATCG GGCTTGGGGC CAGGTCACCA CGTCTGATC 3710
 GGAAGAAACA AGTGCCACG TCTGGCCGGG CGAGGGCACA TTGGTGGCTC ATGCCCTCTC CTCTGCCGG 3780
 AGGTCTCTAC CTGACAGACG CTCCAGCGCT ACATGGCACA TTGGTGGCTC CACCTGCGAGG AGACCAAGGCC 3850
 GCTGAGGGAT CGCTGGCTCA TCGAGCAGGT CTGGGCACTG CCTGGCAGGG TTGGGCAGGG ACTCCACAGCA 3920
 55 GTGGGCTCTC CCCTGGGAA TCACTGGCT CATGACCGGA CAGACTGTG TGCCCTGGGG GCAGTGGGGG 3990
 GAATGAGATG TGATGGGGC ATGATGACCT GTGTCCTG GGGAAATCTC AGCTGGGCCA TGCCAGGCTG 4060
 CGACAGCTG TGCTTACCTG CACCTGCTCA CGTTTGTACTG CGGGGGCTCT CTCAGTCCC GCAGTGGCTT 4130
 TGTTCATGAT TTGCTTAATG TCTCTCTGC CAGTTTGTAT CTTGAGGCCA AAGGAAAGGT GTCCCCCTCC 4200
 TTTAGGAGGG CAGGGCATGT TTGAGGGCTG TCTCTGCCAG ATGCTGGGT CTGAGGGCCAA 4270
 60 AGGAACATG TCCCCCTCT TAGGAGGACG GGGGGTGTGTT GAGCCACGCC CGCTGAGGC GGCTCTCA 4340
 TGCTGGGTCT GTCCACGTGG CCCTGTGGCC CTTTGGCAGAT GTGCTGTGTC CACGTGGGCC TGTGGCTCTT 4410
 TGCAAGTCCC TGTTAGCACT TGCTGGCTC TAGGGGACAG TGCTGTCCAC CGCATGAGGC TCAGAGACCT 4480
 CTGGGGAAAT TTCTCTGGCT CGGGGGAGGT GGCTGGGGCT GCTGGGACCC AGACCTGTG 4550
 CCCGGCAGCT GGGCAGCAAC TCTGGATCA CATATGCCAT CGGGGCCACG GTGGGCTGTG TGGGTGTGAG 4620
 CCCAGCTGGA CCCACAGGTG GCCCAGAGGA GACGTTCTGT GTCACACACT TGCTGCTAAGC CCATGTGTG 4690
 65 CTGGAGAGAC TCGGGCCGGG CAGGCCAGCA TGCCCTGCA TTCCAGGCCA GGGGGCACT TCACTACAAA 4760
 CACTGACCCC AAAAGGGGACG GAGGGTCTTG GCCACGTGGT CCTGGCTGTC TCAGCACCA CGGGCTCACT 4830
 CCCATGTGTC TCCCCTCTGC TTGGCAGAG CTCTCTCCCTG AATGAGGCCA CGAGTGGCCCT TTGAGGCTG 4900
 TTCTTACGCT TCAATGTCGA CCACGGCTG CGCAGTCAAGG GCAAGTGTGAGT CAGGGGCCA GGTGGCATTG 4970
 70 CCCCTGGGGT GGCTGGGGGG GCTGGCAGGG CCTCTCTCTC CCTCTCTCTC GCCCCCTCCC CACTGNCCCT 5040

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCG	AGCCCGGAGC	ACACGAGGT	GTGACATT	AAATCCACTA	10500
	AGATTCACTC	GGGGGGAGCC	CAGGTCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGAC	10570
	AGAGGAGACG	GGGAACGCTG	CTTCTGTG	GCAAGTTCT	GAGGGTGTG	GCCAGGGAGG	TGGCTCAGAG	10640
5	TGTATGTTGG	GGTCCCACCC	GGGGCAGAAC	TCTGTCTG	ATGAGTCGGC	AGCCATGTA	CAGGAAGGGG	10710
	TGGCCACAGG	GAGCTGGAA	TGCACCAAGG	GAGCTGGCA	GCTGCCAGG	GTCAGGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCG	GGGCCCCAGG	AGAGGCGGCA	GGAAAGGGAA	GGGATGCCA	GGCCAGAGCA	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTG	GCTGGGTGAG	CGAGGCTCAT	GACTGGCGA	GGGAACCTCC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTGTC	AGCTCACAGC	CCAGCGAGGT	CCCGCGCTC	AGCAGGAAC	10990
	CAGAACCCCTC	CCCTTGTG	AAAGCACACG	AGATGCCCTA	GGAGAAAACA	GGCAAAGTCG	11060	
10	TTGAGAACAG	TCTTAAAGA	AGGTGGGATG	GTGGCAATT	TTTGTCTG	CCCCGGGACCA	11130	
	CAGATGAGTC	TATAACGGG	TTGTGGTGT	GCCATGGGG	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCCTGGCT	CAGGGCAGG	TGAGTCATGC	TCTGCTACCT	11270	
	GTCTGCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTG	CGAGGGCTGG	GTCCAGGCTC	CTCAGAGCTC	11340
	CTGCCAGGCC	CAGCACCCCTG	CTCCAAATCA	CCACTCTCT	GGGTTTCTC	AAAGCATT	ACAAGGGTGT	11410
15	CAGGTTACCTC	CTCTGGTAC	GGCCCGCCAT	CTCTGGGCTG	ACATCTCTG	TCTGCTTAG	GACCTGTC	11480
	CGAGGTGTC	CTGAGTATGG	CTGGCTGGT	AACCTGGGA	AGACAGTGTG	GAACCTCCCT	GTAGAAAGAC	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCCTATT	CCCTGTTG	GGCTGCTGCT	11620
	GGATACCCCG	ACCCCTGGAG	TGCAAGAGCG	CTACTCCAG	TGAGGCACCC	TGAGGGAA	TGGAGGCTGT	11690
	GCCCGGTG	GGCAGGTGCT	CTGGCAGGGC	CGTGGCTGCC	ACCTCTGTT	CGGTGTTGGG	CAGGGGACTG	11760
20	CCAATCCAA	AGGGTCAGAG	GCCACAGGGT	GGCCCTCGTC	CCATCTGGGG	CTGAGCAGAA	ATGCATCTT	11830
	CTGTGGGAGT	GAGGGTGTG	ACAAAGGGG	CAGTTTCTG	TGCTATTG	TTAAAGGAA	ATGGTGAC	11900
	AGACCTGGGT	GCAGCTGGAGT	GTCTTCAGAA	AGACAGTGTG	ATCCGAACCC	AGACCCCCG	GGCCCTGTC	11970
	GGCGTAGTC	TCTCAAAACCC	GAACACAGG	GCCCTGTCG	GCATGAGTC	CTCTGAACCC	GAGACCCCTG	12040
	GGCCCTGCTG	TGCGTAGTC	CTCCGAACAG	CAGAGACTCT	AGGGCCCTG	TGGGGTGTGAG	TCTCTCCCTG	12110
25	GTGAGCCCCA	CACTCCAAAGG	CTCATCAC	GTCTCACAGG	TGCCATGAGT	TATGATCAC	GTGACCA	12180
	TCAGGGGACA	GGGCCATGTT	GTGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTC	CAGAGCCGA	12250
	GAGCTCAGG	CCCCGTCTA	GGCTCAGAG	CAAAATGATT	GAAGATGCA	ACAGATGCA	AAATCTGTC	12320
	TGTTTCTTT	ATGAATAAA	ATGATCAACA	TTCCAGGAG	GGCAAGGTTG	CTCACCTA	TAATCCCGC	12390
30	ACTTTGGGAG	GCCGAGGTG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGG	CAACCTAA	ACATAGTGA	12460
	AATTCCATT	CTACTTAAA	AATACAAA	TTAGCTG	CTGAGGTC	ACGCGCTGAC	TCCCCGCTAT	12530
	GGCGGAGGCT	GAGGCAAGG	ATACATTG	ACCCAGGAG	CAAGAGTTG	AGTGAAGGCA	GATCACACCA	12600
	CTGCACTCCA	GGCTGGCAA	CAGAGTGAGA	CTTCTATCTA	AAAAA	AAAAAGTATC	AGCRTTCCAA	12670
	AACCATAGT	GACAGGTGTT	TTTATCT	GTCTCTGAT	AAATTTACT	GTTGCTGTC	TAGAGGCCCCG	12740
35	AACTGGGGT	GGCCTCTCT	GAAGGACACA	CCTCTGAGG	AGAGAAATA	AGTGGTGA	GGTTGTTAA	12810
	CCAGAGGTTT	AAACTGGGG	CTCTGCTG	TGAGTTAAC	GTCCAGATG	GGACTTTG	TCTTCCAGA	12880
	ATGCTCCCTG	GGGTTGCTT	CATGGGGGAG	CAGCAGGTT	GGACACCCCT	GTGATGGGG	AGCAGCAGGT	12950
	GCAGACGCC	TCATGATGGG	GGAGTGGCAG	GTGAGCACAT	CCTCTGCT	GGTGGCCAGC	ATGCTCCCTG	13020
	TGCAAGTCTC	TCCCCACAA	GATGCGGTC	TCTGCTGCTC	CCACAGCTC	CTGCTTCCCT	CTCACAGCCT	13090
40	TACCTGTC	GGCCCTCTG	TGCTCTGTC	TGATGATT	CCACATTTC	TGGGCTCCA	GCACCTCTTC	13160
	GCCTCTCCA	GGCACCTCTG	CACTGCTG	CATACAGTC	AGCTGTGAC	TGTCCTAGC	TTATTTCTG	13230
	CCCCCATGAA	TGTTTCTT	AGGACAGGG	CCCCCTGTTG	CACAGCTTGG	ACAGCATCA	GTGAATGTTA	13300
	TTGAAGGACA	AAAGACAGAC	AAACAAATCA	GGAAAATGGG	TTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
	GGCTAGTGC	GGATGGGTGG	GCATCAGGTC	ATCAGATG	GGTCCAATG	CAGAATATT	TGTCCTCCA	13440
	AAGGGCAACTT	GGTCAGAGTG	TGTCCTGTC	GAGGGTGGTC	TAAAAGCTCA	GCAGTGGAGG	CASTGGTTCG	13510
45	CCATACATCG	GGTGAACCTA	CATCTCTG	GTCTGAAGT	TACAGCAGG	GCTTGAAGGG	CATCTGGAG	13580
	AAGAAAACAG	GCACAAATGAT	TAAGAAGAGT	GGAAAAGGAA	AGTGGTAAG	ATGGGAATT	TCTGTCCAG	13650
	ATTTTACTCT	CCCCAACAC	AGCTCAGATG	GTAGAATG	GTGAGCAACTG	ATGGAGCTG	AAATAGAAC	13720
	AAACGGAAAC	CCTATCTCTC	AAACAACTGT	TTTAAATGTTG	TATGTTGGC	AGCTGATGGA	AAAGAGAGTG	13790
50	TGTGTGTAAT	TTTTTTCT	GAGAAAAC	ACTGGAAAGCA	ATAAAGTGT	GTCTTACAG	CATATACCA	13860
	AGCAGATCT	AGGTAGAAGA	GGAGACACAT	GCACAAACAA	CCAGCAACAA	AAATAAAACA	AAAGACTCAA	13930
	AGGGGAAGGGA	GGTGAACGTT	CCCTGGTTG	GTGTTGGG	AGGACACACA	GGGAGGCGG	TGAACACAGT	14000
	GAGGCAACGG	GCATTGCTT	CACTGCAGAG	AAACTCAGT	TGCTGAGG	ACAGTAAAA	TGGCCATTCC	14070
	CTGAGGCTT	TGTGACGCTG	TTTATTTA	GGGGCCCTG	GAGGCTCTGG	ACATTCTAC	TCTCCTTTG	14140
55	TTTCCTAAC	CACCTGAGAG	TGAGGAGGAG	AAAGGCTC	GGGGAGCAGC	GGGCGTGTG	CACCCAGCTG	14210
	GCAAAAGGCA	TGCACTGATT	CAGCCTGGC	TCTGCTCCG	GGGGCGTTCG	TCTGCCGAG	GACCCACAC	14280
	AACTGACACC	CATAGGCTCA	GGGTGAGCC	GAGGGCTCA	TCTGTTGGG	GATGGCTG	AAAGAAGAAA	14350
	TGGACGCTG	ATGACACACTT	GGGAAGGTC	TACCGAGC	TGCAAGAAA	TGATGTA	ACTGACAGCG	14420
	AGCCCCATCC	CTCAAAAGAA	CGCACGTGAA	ACTGATGGC	AGACCTGTC	CCATCCCTA	TGCTGGCTC	14490
60	TTTCTGGG	TTGCAAGAG	CCAGCAGTC	TGTTGAGGAGA	GCTGAAAAGA	CTTCTGGA	AAGCAGCTTG	14560
	TTTGCATGGA	AGTCTCTACA	ATGCTCTG	TCTTCCTG	AATTCTCA	CTGAAGTGC	CACAGATTAT	14630
	CACGGGCTT	TTTACGATT	TCCAGTGTTC	CAGGGCAGGG	GACTTGCAC	AGCAAGTCA	GAACCTGCC	14700
	AAATACAGG	CTAAGGAGAT	ATTATGCA	ACAAAAC	CTCTGCCAT	AAACATT	AAAGAATT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAAC	TTTCACTG	AGCAGTGTG	AAAGCTGGAT	GTAAAGAAC	14840
65	ACACCCAGG	AGGCTGCCGT	GAATGTCATG	TGTTGTCATC	TTTGACATG	GACATACATG	GGCAGTGGAT	14910
	GGTGGTGGAGG	CCCTGGAGGA	CATCGGTGG	ATGCTCCAT	CCTGCCCCAT	TGAGGACACC	ATGTTGCCA	14980
	CGTGCACCTCA	CTGGGACCC	TTTACGCTG	TGCCCCATTG	CTCTGCCATC	CTTGAGATT	AAACACAGT	15050
	AGATCCCCCA	CGCCCAACTC	AGTGTCTCC	CACAAAAAAC	CTGACTCACA	CCTGTTCTCA	CTCAGGGGAC	15120
	GCCCCGGAGC	GCCCCGCTCA	CAGTTTATA	TGTTGTTTTG	GCTGAGTTAT	GTGAGATCT	CATCAGGGCA	15190
70	GATGATGAGT	GCACAAACAC	GGCCCTGCGA	GGTTGGATA	CACTCACAT	CACAGGCCAG	TGCTGGTGG	15260
	AGTTTGGTCA	TGCAAGAGTC	TGAGTGGT	AGTGGTGTG	GTGAGCACCC	AGCCCCCTCG	15330	
	GGCTGCACCG	CATGCCCTAC	GCAGGACAG	GAAGCGGGAG	GAAGGAGGA	GGCTCTTTG	AGCAAGCTT	15400
	GCAGGAGGGG	GCTGGGTG	GGGCAGGCAC	CTGTCATG	CATTCCCCC	TGTTGTCAG	CTATGCCCGG	15470
	ACCTCCATCA	GAGCCAGTCT	CACCTTCAC	GGGGCTCTCA	AGGCTGGGAG	GAACATGCGT	CGCAAACACTT	15540
75	TTGGGGCTT	GGGGCTGAAG	TGTCACAGC	TGTTCTGGA	TTTGAGG	ACGAGGCTGA	TGGTCAGCAC	15610
	AGAGTTCA	GTTCAAGGAGG	TGTTGTCGCA	AGTATGTTG	TGTTGTCG	CGCGCGCTG	TGCAAGGCT	15680
	ATGGTACTG	GCTGCACGTA	AGAGTCAC	TGACAGCATA	TACACGTGAG	CACATACATG	TGTGCATG	15750
	TGTACATGAA	GGCATGGCAG	TGTTGTCAC	GGTGTGCAAG	GGCACAAGT	TGTCAGCATG	CGAACATGCA	15820

5	CCTGACATGC ATGTGTGTTC GTGCACAGTC GTGTGGGCAT TCACGTGAGG TGCATGCGTG TGGGTGTGCA 15890
	GTGTGAGTAG CATGTGTGCA CATAACATGT ATTGAGGGT CCTCGTGTTC ACCCGCTAG GTCTCAGCA 15960
	CCAGTGCAC TCTTACAGG ATGAGACGGG GTCCCAGGC TTGGTGGGCT GAGGCTCTGA AGCTGCAGCC 16030
	CTGAGGGCAT TGTCCCACCTT GGCGCATCCGC TCTCGACTCC TCTCGTGTG GCTTCTGTG TCAACTCCCC 16100
	TCTCCGTGTC GCATTTACAT CCACCTCACT CCTCTCTCC TGTTGGCATC CGCGTCACT CCCCCCTCT 16170
	GTGGGCATCT GCGTCCACCT CCCCTCTCTG TGGGCATTT CGTCCACTCC CTCTCTGTG TCTTCTGT 16240
	CTTGGCGGAG CTCGGGGGCG AGGAGAGATGA AGACAGTCT TTCTTGCAA ATATGTCCTC TTCTGTGTC 16310
	GTGTAGGGCC AGGGCGGATT TCACTGGAA GAGGGATAGT TTCTTGCAA ATATGTCCTC TTCTGTGTC 16380
10	CATCTGAATG GATGATAAAG CAAAAGTAA AACCTTAAA TCCAGAGAG GTTCTACCG TTCTCACTC 16450
	TTTCTTGGGC ACTCTAGGTG AGACAGTGTG CACCAACATC TACAAGATC TCTCTGTGCA 16520
	GGCGTACAGG TGAGCCGCC CAAAGGGGTG CAGGGCCAGC CCTCAGGGAC CCTCCCGCTG CTGCTACACT 16590
	CTGACCCGGG GCTTCACCTT GGAACTCTG GTTTTAGGG GCAAGGAATG TCTTACGTT TCAAGTGTG 16660
	TCGCTGCTGT GCACAGTCTT GTCTGTGCA AGACAGTCTT CTCCATCTCT GGGTAGTGGT 16730
	AGGAGGGCTG GTGTCGGCAG TGTCCTCCAC TGTGCTGTG CACTGGCGT GGAGCGTCAT GGAGGCCATC 16800
15	CCAGGGCAGC AGGGGCATGG GTTAAAGAGA TGTGTTAGGG GAGTCTTAGC AGAGGAGGCT GGGAAAGGTG 16870
	CTGAACAGTA GATGAGGAGAT CAGATGGCCG GAGGATTTGGG GTGTCAGCA AGAGGGCCG AGGTGGGTG 16940
	ACGGTGGAGGT CGCTGGCCAG ACCCCCCGGG AGGTGAGCGA GAGCTGTGCG TCCCCACACA GCCCCGGCAG 17010
	CACCTGTGCT CTGGGCATGG CTGTCCTCTT GGAACTCTG CGTCTCTGC TGTCAGGGG GTGCCCTG 17080
	CAAGAACATG CAACCTTATC ACAGAGGGAA GGCGCAATCT GTGTCAGCA CAGGGCCAGC TTCTGCTG 17150
20	AGTCAGGGCA GGTGGTGGCA CAGGGCTGG AGGCTGTACCA AAGGGCAGTC GGGCACCAACAGC GGGCCGGGCC 17220
	TCCACCTCAA CAGGGCTCCC GAGGCACTGG GAGCTGAATG CCAGGAGGCC GAGGCGCTCG CCCCATGAGG 17290
	GCTGAGAAGG AGTGTGAGCA TTGCTGTGCA CCAGGGCTGG AGCTGTGCG ATTACCGTGC ACACTTGATG 17360
	TGAAATGGAG TCGTCGCTCA CGCAGCAAGGG CTACAGGGAG ATTTTCCAT TACAAGGTG 17430
	TACCATGGAA ATGGTTTTA ACCCGAGTGC TTGCGCTTC ATGCTGTGCA AGGGAGGGCA GAGCCACAGC 17500
25	TGCATGTTAAC CGCCCTTGCA CCAGCTTACAG AGGCTTGGGA CCAGGCTGTC AGCTTCCAG GAGTGGTCCG 17570
	GCTCAGACCG CCCTCTCTC TGCTCTCTC TGCTGCTCA AATCTTCCCCT GCTTGTGATC TCCCTGACGC 17640
	GTGCCATGGC CTCCTGCAA CTGCTTGTGAC TCTCTTCCGG AAACCCCTGG GGTGTCCTGG ATACAGGTG 17710
	CACTGAGGAC TGGAGGTGTC TGACACTGTG GTGACCCCA GGTGTCCTGG GGGCTCTT GGGCCCTCTT 17780
	GGGCCATGAT GAGGTCAAGAG GAGTTTCCC AGGTGAAAC TCCCTGGAAA CCTCCAGGGC CATGTGACCT 17850
30	GCCACCTGCT CTCCTCATGG TCACTGTCACT TTGTCCTCA TTTCCCAACCC AGGGTCTCTA GCTCCGAGGA 17920
	GCTCCCTGAG AGGGCTCTGG CTCAGGGCG AGGGCTGAG TTTCCCAACCC CATGTGGGGA CCCTTGGGA 17990
	GTGCGTTGAT TGGTAGGCC TGAGGAGGCC GAGATGGAT CCGCCACAGG CGTTTCCAA ACACAGAGT 18060
	AGGCACATGG AGGGCCAGG CTCCTGAGGCA GAGCTGGGAG AACGAGAGGC TGGGCCCGA 18130
35	TTTCACCGCA GCCAGGCTGC AGTGGGCGAG GGTGTTGAG TCCACGTCG GCTGGGGCG GGTGTCGATT 18200
	CAAATCCGCT GGGGCTCGGC CTTCCTGGCC CGTGTGGCC GGGCCTCCAC AGGGGCTTGG GGTGAGGCC 18270
	CCGACCTCTA CGAGGCTGGCT ATTTCTCTT TTGGAAGAGA GGGCCCTACCC CATGCTAGGT GTTCCCTCTC 18340
	TGGGTCAAGA GCGTGGCCGGT CGGCAACCC CGGGACCTTA CGGTTATTAA AACATTCTG 18410
	GGCTGGCTTC CGTTGTGCTT AAATGGGAA AAGACATCCC ACCTCAGCAG AGTTACTGAG AGGCTGAAAC 18480
40	CGGGGTGCTG CTTTGAATGG TGTGATCTCA GGTGTCATCCA GAGTGGGAG AACGAGTCAG TGGAGCAGG 18550
	TACATGGGGG CTCAGGCGAG TGAGGAGATGA GAGTGTACAG GGGGGCTCAG CGAGTGGGAG AGGCCAGGT 18620
	CATGGGGGGC TCAGGCACGT GGTGAGATGA GGTACACGGG GGGCTCAGGC AGAGGGTCAG ACCAGGTACA 18690
	CGGGGGCTCT GATCACACGGC ACATATGAGC ATGATGTCAC ATGTCATGGT ACAGTGTG 18760
	CACACCTGCC CAAAGTCCC AGGAAGCTGA GAGGCAAAAG ATGGAGGTG AGGGGCTGG CGGGGTGCTG 18830
45	CACACCTGTA GTCCCAAGCAC TTGGGAGGC CGAGGGCGAGA GGATCCCTT AGCCCAGGAG TTAAGACCA 18900
	GGCTGAGCAA CATAGTAAAGA CCCCATCTCT ATGAAAATAA AAAACAAAAA TTAGCTGAAAC ATGTTGGT 18970
	GGCCCTGTAG TTCAAAACTA TGCGGAGCTG AAGTGGGAGG ATCACTTGTG CCCCAGGAGT GGAAGCTGCA 19040
	GTGAGCTGAG ATTGACACAC TGTACTGTG CCGTGGTGCAG AGACTGAGAG CCCATCTCA AAACAACAAA 19110
	GAAGACTGAC TCTTGGAAAG AACACATTAG TAGGACATCA ACCTACACAG AGAAAGCCAAG 19180
50	TCCGGTCTC CGGTGTGAGT AGATGAGATG ATGGTCTCTC ACACCATCAC CCCAGACCCA GGTTTATG 19250
	ACACACAGGG CGGGTGGCTC AGAAGGGATG CGCAGGACGT TGATATACGA TGACATCAAG TTGTCGTCAC 19320
	GAAGGGCAGG ATTACATGATA AGTACCTGCTG GTGACACAGG GAACATGGAA TAAACTGGAA ACCTTAGAGG 19390
	CCTCTCCCGA ACAGGGGCTA ATCGAACGGG AGCATGGGG CCTGGCATC AGGATGGAGC TGCTTCAGCC 19460
	TCCACATGCC TGTTCATACA GTGGTGCAC AGAAACGAG TGTACCTGTG CACACACAGA CACGCACCTA 19530
	CTCGCACACA CAGCACACAGT CGACAGATCG ATCGATGCA TCCGGTGTG TGCACTGTG CCCATGAGGA 19600
55	AACCCATGCC TGTGCAATTCA TGACCCGACA CGGGCACCCG TGGGGCATG CCTACACCCA CGAGCACCGT 19670
	CTGATTAGGA CGCCCTTCTC CTGACCGTGT CGGCCATCT CCTAGGTGTT ACGGCATGTG GCTGCAGCTC 19740
	CCATTTCATC AGCAAGTTG AGAACACCC ACATTTTCC TGCGCTCAT CTCCTGACCC GCTCCCTCT 19810
	GCTACTCCAT CCTGAAGAGC AAAGAACCGAG TGTATGTCAG GTCGGCTGCC TGAGTGGCAG CAGTGCCTG 19880
	CTGCTGGTGT TAGTGTGTCAG GAGGACTGAG TGAATCTGGG CTAGGAGAT TCTTACCCCT TTGTCATCA 19950
60	GGAAAGTGTG TTACCCAAAC CTGTCAGGCC TCGTCAGGCC GGGGTGAGCA GAGCACCTGA 20020
	TGGGAGGGAC AGGAGCTGTC TGAGGAGCTG CCTCTTCCC ACCTTGTCTG GCTCTGGGA CCCTGGGG 20090
	GCCCTGGTCTC TCTGTGTTGC CCCATGGTGG GATTGGGGG GCGCTGGCTC TCTGTGTTGC CCGTGTG 20160
	GATGGGGCTG CCTCCCGTCA ATGGCACTT GGGCCCTGTG GCAARCCCCG GCGAACGGCT TAGGAGGAGG 20230
65	CCAGGGCCAG CCTACCCAC CCTCTCTCAGG AGCAGAGGCC GCGTATCAC ACAGCACAGC CCCGGCGCT 20300
	CCTCTGCTTC CCAAGTCACCG TCTCTGCCC CTGGACACTT TGTCAGCAT GAGGGAGGTT TCTGATCCGT 20370
	CTGAAATTCA AGCCATGTGCG AACCTGGGT CTCAGGCTTA AGCGATCTCA CTTCTGTTC TTCTGTGTT 20440
	GTGGAATTTC CACCTGGAGA AGCCGAAGAA AACATTCTG TCAGTGTACTC TGCGGTGCTT GGGTCGGAC 20510
	AGCCAGAGAT GGAGGCCACCC CGCAGACGGT CGGGGTGTGGG CAGGTTTCCG GTGTCCTCTG GGAGGGAGC 20580
70	TGGGCTGGGC CTGTGACTCC TCAGCTCTG TTTTCCCCA GGGATGTCCC TGGGGGCCAA GGGGGCCGCC 20650
	GGCCCTGTG CCTCCGAGGC CGTGCAGTG CTGTCGACC AAGCATCTCT GCTCAAGCTG ACTCGACAC 20720
	GTGTCACCTA CGTGCACACTC CTGGGGTCAC TCAGGACAGG CAAGTGTGGG TGAGGGCCAG TGGGGGCC 20790
	ACCTGCCAG GGGTCATCT TGACGCCCT GTGTCGGGG AGCAGCTCA ATGTCGCTG AAGTGCAGAC 20860
	GGCCCCGGGC CTGACCCCTGG GGGCCTGGAG CCACGCTGGC AGCCCTATGT GATTAACAGC TGGGTCCCC 20930
75	AGGCCACAGGA AGGGCCACCTG TCTTGAACCC CTGCTTCCCA TCTTCAGGGG GATGGCTCCC 21000
	CACGCTTGGG AGCCCTTGTG CCCCCGACCT GTGTCCTCTC ACAGCTCTT CCCCCTGTC TGCCCTGAC 21070
	TCCTGGGGTC CTGAGCAAGT TCTCTCCCG CCCCCCGCT CCAGCGTCAC TGGGTGCTCC GTCTGCTGCC 21140
	CCCGGTGGAG GGGGTGTCTGT CCCTTCACG AGGTTCCAC CAGCCAGGGC CAGCGAGGTG AGGGCCCTGCC 21210

5	TGCCCGGCA CCCCACAGTC CTAGGAGGTG TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCG 21280 ATTTGGCC CGCAGCCAG ACGCAGCTGA GTCGGAAGCT CCCGGGAGC ACGCTGACTG CCCTGGAGG 21350 CGCAGCAAC CGGCACACTGC CCTCAAGACTT CAAGACCATC CTGGACTGAT GGCCACCCCC CAACAGCAG 21420 GCCGAGAGCA GACACCAGCA GCCCTGTCA GTCTGAGACTT ACCTCCAGG GAGGGAGGGG CGGCCACAC 21490 CCAGGCCGC ACCGCTGGGA GTCTGAGGCC TGAGTGTAGT TTTGGCCAG GCCTGCATGT CCCGCTGAAG 21560 GCTGAGTGTG CCGCTGAGGC CTGAGCGAGT GTCCAGGCAA GGCGCTGAGT TCCAGCACAC CTGGCTCTT 21630 CACTTCCCCA CAGGGCTGGC CCCCTGTCA GTCTGAGGCC AGCTTTCTT CACAGGAGC CGGGCTCCA 21700 CTCCCCACAT AGGAATAGTC CATCCCCAGA TTGCGCATTTG TTCAACCCCTT CCCCTGCCCT CTTTGCCTT 21770 10 15 20 25 30 35 40 45 50 55	CCACCCACAT CATCCAGGTG GAGACCCCTGA GAAGGACCCCTT GGAGGCTCTG GGAATTGGA GTGACCAAAG 21840 GTGTGCGCTG TACACAGGGG AGGACGGGAGTGG GGGTGTAGTGG GGGTCTCTTGGG GGGGGAGG 21910 GCTGTGGAG TAAATACTG AATATATGAG TTTTCACTGTTT TGAAGAAAAA TCTCATGTTT GAATCTTAT 21980 GTGCACTGCA TAGACACCAC TGATGCAAT TACAGAARGCC TGAGTGTGAA CGGGGTGGTG GTCACTGCGG 22050 GCCCATGGCC TGGCTGTGCA TTACCGGA TGCTATGAGT AATGGGGTGG TGCTGAGTCC CGGGCCATGG 22120 CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGGAGG AGGGAGGAGA GGGTAGGGGA TAGACAGTGG 22190 GAGCCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGCC GAAGGGCAGC AGGGATGCTG GGGGCCAGC 22260 TGGGGCGGC GGGATGATGG CAGGGCTGGC CAGGGCTGGC GGGATGATGG GGGGCCACCG TGGGGTGGCA 22330 GGGGGTATGG GGGGGGCTGG TCTGGGTGGC GGGGAATGAGT GGGAGGCTG GCTGGCCCCC CTCTCCCTT 22400 GCCTCCACC TGCAAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTCTAT 22470 GGAGGGGGGG GGCAGGGGCA TGACACCATC CTGTAAGAAA TCCAGGATCTT CTCCTCTGCA AGGCCCCAAC 22540 TCAGGTGAA AGTCACATTC CGCCCTCTGG CATTCTCTTA AGAGTACAGC AGGATTCTGA TCTCTGAAGG 22610 GTGGTAGGG TGGGGCAGTG GAGGGTGTGG ACACAGGAGG CTTCAAGGGT GGGCTGGTGA TGCTCTCTA 22680 TCTCTTATC ATCTCCAGT CTCACTCTC ATCTCTTAT CTCATCTCTG TCTCTCTTA 22750 TCTCCAGTC TCATCTGCA TCCCTTACCT ATCTCCAGT CTCACTCTT ATCTCTTAT CTCTCTACT 22820 CATCCAGACT TACCTCCAG GGCCTGGTGC AGGCTCCAG TGAGGCTGGA CATACTGCTT CTCTCAGGCA 22890 GAAGGAATCTG GAAGGATTCG AGAGAACAGG AGGGGGGGCT CAGAGGAGC CAGTCTGGGG TGAGAAGAAC 22960 AGCCCCCTCC CAGAGTGG CTTGGGCCAC AGGAAACCGA GGGCCCTGGC TGAGTGGCTC CAGAGCTTCTC 23030 CAGCAGTCTC CTGGTGGGGC CTTATGGTAT GGGCGGGTGC TACTGAGTGC ACCTGGACCA GGGCTTCTGG 23100 TTTGAGTGC GCCCGGAGCT GCGCTGGTGC GGGGGGGGGC CTATGGCCA CTGGATATGG CGTCATTAT 23170 TGCTGCTGCT TCAGAGATG TGAGTGTAC CGAGCTTAAT TGAGTGTGGT GGGCCAACTG CACAGACTG 23240 GTCGAAATG CACTCTGGTG CCTGGAGGCC CGGTATAGGA GCTGTGAGGA AGGAGGGCTT CTGGCAGGCC 23310 GGCCTGGGGG CGCCCTGGCC CTGCAAACGT GAGGGAGGC GCCCCGGGGC CGCTGGGGGG AGCACCTCAA 23380 GTGAGGGTT GGAGACACAA GGGCGGGAC TCTCCAGGAG CAGAGGGCCG TGCTCAGGCA CACCTGGGGT 23450 TGAATCACAG ACCAACAGT CAGGGCATTC TTCACTTAC AAAGCTCCAG ATTCCTGTTT 23520 35 CTCGGGGTGT TTGTTGTA ARTTTACTC AGGATTACTT ATATTTTTG CTAAAGTATT AGACCTTAA 23590 AAAAGGTATT TGCTTGTATA TGCTCTTAACCT ACAGACACAA CTACTTATT TGCTGTGTTT TATTTTATT 23660 TATTATTATT ATTAGAGATG TGCTGCTACTC TGTCACCCAG GTGTTTAGTG CAGTGGCACA GTCATGGCTC 23730 GCTGTAGCCG CAAACCCCA CGCTCAAGTG ATCCTGGG CTCAGCTTCC CAGAGTGTG GGATTACAGG 23800 TGTGAGCCAC TGCCCTTGCC TGGCACTTTT AAAACCAACT TGTAAAGTC AGGTCAGTG GCTTCCACAC 23870 CTGTCTACCC AGTAGTTGGG GAAGCCGAGG CAGAAGGTT GTCTGAGGCC AGGAGTTGGA GACCAGCATG 23940 40 GTTAACATAG GGAGACCCCA TCTCTACAA AAATGAAAAA AGTTATCGG GCGTGGGGTC CAGCATCTGT 24010 AGTCCCAGCT GCTCGGGAGG CTGAGTGGGA GGATCGCTTG AGGGCCGGGG GTCATGGCTG CAGTGGCTG 24080 TGATGCTACC ATCCGACTCC AGCCCTGGCA ACAGATGAGC ACCTCTGCTC AAAAAGAAAAA AAAAAGAAG 24150 AAGGAGAGG AAGAAGAGAAG AAGAAGAGG AAGAAGAGG AAGAAGGAAG AAGAAGGAG 24220 AAGGAGGCCCT GCTAGGTGCT AGGTAGACTG TCAAATCTCA GAGCAAAATG AAAATAACAA AGTTTAAAG 24290 50 GGAAAGAAAAC ACCCCAGCTC TTGGACCTTC CTTAGGCCCTG AACTTCATCT CAAAGCAGCTT CTTCCACAG 24360 ACAAGCGTGT ATGGAGCGAG TGAGTCAA CGAGAAAGGG AGGAGAACCA GGCAGGGTG GAGGCTGTGG 24430 GTGACACCAG CCAGGACCCC TGAAAGGGAG TGGTTGTTT CTCGCTCTG CCCCCACGCTC CTGGGGTCC 24500 TGCACCTGCT GTAACCGTCC ATGTTGGTCC CAGGTGCCCA CCTGGGAAGG ATGCTGTGCA GGGGGCTTGC 24570 CAAACTTGG TGGTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAAATG ACAGCCCTC CCCAAAGATG 24640 CCCACGCTCT TCTCTGGAA CCTGTGAATG TGTCACTGGC AAGGCAGAGG CTGGTGAAGG CTGCAGGTGG 24710 AATCACGCTC GCGACTCAGC CGATCTTAAG GTCACTCTGG ATTATCTGGT GGGCTGTATA TGCCACAAAG 24780 GCTCCCTAGA AGTGGAGAGG GGAGGGAGGG AGAGGTCAA GAGGGGGAGT GAGAAGGACCT ACTGGCCACT 24850 GCTGGCTTG AGATGGAGGA GGGGGTCCCC AGCCAAGGA TGGGGGAGC CGCTCCATGC TGAAAAGCA 24920 AGCAATCTC CCCGGCTCTG AGGGCACACGG GCGCTGCCCA CGCTCCAGT TGCAAGCTTGC 24990 55 TCAGCTTCC CGGCCCTCCAGA CTGTAAAGT GATGCGTTTG TGTTCAGGCA CTAAGCTGCA GTGATTCTGTC 25060 ACAGCAGCAA ATGGAATAGC AGTACAGGGAA ATGAAATACA GGGACAGTTC TCAGAGTGC TCTCAGGCCA 25130 CCCCCTGGG 25138
---	---	---

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon					Intron					3'-Exon					
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk. (%)	70	60	80	100	100	95	70			80	100	100	60		

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);
Intron 2: bp 12951-21566 (SEQ ID NO 5);
Intron 3: bp 21763-23851 (SEQ ID NO 6);
5 Intron 4: bp 24033-24719 (SEQ ID NO 7);
Intron 5: bp 24900-25393 (SEQ ID NO 8);
5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position

21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCGGGTCGGCTCCGGTGGGTTGAGGGCGCCGGGGAAACAGCGACATGCGAGAGCAGCGCAGG
CGACTCAGGGCGTCTCCCCCGAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCCAGAGCTGAATGCAGTAGGGCTCAGAAAAGGGGCAGGCAGAGCC
CTGGTCCTCCTGTCTCCATCGTCACGTGGCACACGTGGCTTCGCTCAGGACGTCAGTGGACACGGTATCTCGCC

10

TCTGCTCTCCCTCTGTCCAGTTGCATAAAACTACAGGAGGTCACCTTACGTTGATGGACACGGCTTCCAGGC
CGAGGCGAGAGCAGTGAACAAGGAGGCTGGCGCGCAGTGGAGCCGGTTGCCGGAATGGGAGAAGTGTGGAAG
CACAGACGCTTGGCGAGGGTGCAGGTTACCTATAATCCTTCGCAATTCAAGGGTGGAAATGAGAGGTTGG
CGAGAACCCCCCTCTCCTGGGGTGGAGGTAAGGTTTGAGGTGACGTGGTCAGCCAATATGCAGGTTGTGTTA
AGAGTTAATTGTGTTGACGCCAGGTGCGGTGCGTAACGCCGTAATCCAGCAGTGGGAAGCTGAGGCAGGTGGA

15

TCACCTGAGGTCAAGGAGTTGAGACCAGCCTGACCAACATGGTAAACCCATCTGTACTAAAAAATACAAAATTAGCTG
GGCATGGTGGTGTGCTGTAATCCAGCTACTTGGAGGCTGAGGAGAATCACTTGAACCCAGGAGGCGAGGC
TGCAGTGAGCTGAGATTGTCATTGACTCCAGCCTGGCGACAAGAGTGAACACTCTGTCTTAAAAAAAGTGT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCATCTTAGGT
GAAGAGGGCACATGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGTGGGTT
ATGGTGTGCTGGCCCTGCCGTGCCCCACCCCTGTTTCTGGATTGATGTTGAGGAACCTCCGCTCCAGCCCCTT

20

TGGCTCCAGTGTCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCAGTTCAACCCCTCCCCACAAACTCCAAAGAC
ATGTAAGACTCCGGCATCGAGACAAGGAGGTGACCTTCTGGGCTCTTTTTCTTTTATGGTGGC
AAAAGTCATAAACATGAGATTGCACTCTAACACCGTTCTGTGACAGTGCAGAATTGCTAACCTGGCGGTGTTA
CAGCAGGTTGCTGAAATGCTGGCTTGGTCACTGGAGTCCCTACCCATGAAACGGCAGCTGCTCACACCTGCT
GGCTCAGGTGACCAGCCAGTCAGATAAGCGTCACTGCAACCCAGTTTCTGTTGTGCTCCAGCTCTCGTGTGAG

25

GAGAGTTGAGTTCTGTGAGGACTCTGCTGTCACTGCTGTTCTGACTTCAGATGAGGTCAACATGCCCCGG
CTTATGCAAGGAGTGAGCGTGGTCCCCGGTGTCCCTGTCAGTGAGGTGAGTGAGGCGTTGCCCGGAGGTGT
GTCACGTGTAGGGTGAGTGAGGCGCCGATCCCCGGGTGTCCCTGTCAGTGAGGTGAGTGAGGCGGTGGT
CCCTGTCACGTGAGGTGAGTGAGGCGCCGATCCCCGGGTGTCCCTGTCAGTGAGGTGAGTGAGGCGGTGGT
GTGTCCTGTCCCCGTGCAAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTCAGTGAGGTGAGTGAGGCGGT
30

30

CCGGGTGCTCCCTCTCAGGTGAGGTGAGTGAGGCGCCGGCCAGGGTGTCCCTGTCAGTGAGGTGAGTGAGG
GTCCCTGGGTGCTCCCTCCAGGTATAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTCAGTGAGGTGAGG
CGCGGCCCCCGGTGTCCCTCTCAGGTGCAAGGGTGAGTGAGGCGCTGTCCCCGGGTGTCCCTGTCAGTGAGG
GAGGCTCTGCCCCAGGTGCTTGGCTTGTCACTTGAGCTGCTCTGCAATGTTGCTCTTCTATAGCCACAGCT
GCGCGGTTGCCATTGCTGGTAGATGGTGAGGCGCAGTGCTGGTCCCAAGCCTATTTCTGATGCTCGCT

35

TCTTGGTCACCTCTCCGTTCCATTGTCACGGGACACGGGACTGCAGGCTCTGCCCTCCGCTGCCAGGCAGTG
CCACAGCTCAGGTGGCTGCTGTGGGCTGGCTGACCCAGTCAGTGCCTGGCTGCAACAGTGCATGCTGCCAATACTCC
TCTCCCAAGCTGTCTCATGCCAGGCTGGACTCTGGGCTGCTGTGCTGCCACGTGTTGCTGGAGACATCC
AGGGTTCTGTGCCCCGTGAGGAAAGCAAGTCACCCAGCCCCCTCACTGTGCTGTTCTCCCAAGCTGCCCTCTGC
TTGGCCCCCTGGGTGGCAACGCTGTGACCTTATTCTGGCACCTGCCGCTATTGCTTAGGCTGGCTCTG
40

CCAGTCGCCCTCACATGGATTGACGTCAGGCCAGGTGGAGTGCTCTGCTCTGCTGAGGACACCACGTG

GAGGGCCGGTGTCTCGCCAGCCTCGTCAGACTCCCTTGGGTCTAGTTGAATTCACTGATTACCTCTGACG
 TTTCTATCTCCTTCTGGTTATTCTTCACTCCCTTCTAGCTTCTAGTTAGTCATGCCCTC
 CCTCTAAGTGTGCCCTACCTGACCCCTGTGTTGATGAAAGTAATCTCAACATCAGCCACTTCAGTGTCTAAA
 ATACTTCAAAGTGTAAACTCTTTAAGTATTCTTATCTGTGATTTCCTTGTGACGCTGTGTTGACGTGA
 5 AATCATTTGATATCAGTGACTTTAAGTATTCTTCTAGTTAGCTTCTGTGATTTCTTGAGCAGTGAAGTTATTGAACACT
 GTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTAAAGTATTCTTATGATTGATTCTAAGTCACTCAGT
 TGTTAGTGGTCTGTATAATACCAATTATTGAAGTTGCGGAGCCTGCTTGTGATCTAGTGTGATGGTCCAG
 AACTGTCCATTGTAATTGACATCTGTCAATAGTGGCATGCATGTCAGTATATCCAGTTAAAGTCCAGTGCA
 10 AAGCTTCTGTCCTCTAGATGCATGAAATTCAAGAAGGAGGCCATAGTCCCTCACCTGGGGATGGTCTGTTCTT
 TCTTCTCGTTGGTAGCATTATGTGAGGCTTGTGAGGTGATGCACGTGGTAGAATTCTTATCTCTGATGAGTGAA
 TCTTTGGAGACTCTATGTCAGTAAATTCTTAAATTGCTCTAGTACTGCAACTGGCTTCTGAGTGCACACTGGCTTCT
 TTTGATTAGTATTCTCTGCTGTCTGTTCTGCCTTAATTATATATATATATATTTTTGAGACA
 15 GAGTCTTGGCTGCGCCAGGGTAGTGCAGTGGTAGTGCACAGGTAGTGTGATCACAGGTAGTGAACCTTACCTCTGGCCTGAGCCG
 CCTCTCACCTCAGCCTCTGAGTAGCTGAACTGCAGACACGCACCGTACACCTGGTAATTCTTAAATTCTCTGG
 ACTAGAGACCCGCTGGTGCACTCTGATTCTCACTGGCTGGTAGTGCACAGTGGTTAACGGATAACCTGATTCTTCTTCTG
 20 GGTGCGCATGTGCGTTCTGCGAGTGTGTTGATCCTCTGCTGGCTCTGGTCACTGGCATTGCTTCTGACCCACCTCTG
 CTTTGCTTAGTGTACCCCTGATCTTTATTGCTGTGTTGCTTTGTTATTGAGACAGTCTCACTCTGTCACCCA
 GGCTGGAGTGAATGGCACAATCTGGCTACTGCAACCTCTGCGCTCTCGGTTCAAGCAGTCTCATTCTCAACCTCA
 TGAGTAGCTGGATTACAGGGCCACCACCGCTGGTAATTCTGTTAGAGATAGGCTTCA
 25 TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTGTGATCTGCGGCTTGGCTCCACAGTGTGGGATTACAGGTGCAA
 GCCACCGTCCCCGGCATACCTGATCTTTAAATGAAGTCTGAAACATGCTACCCCTGCTGAGCAATAAGACCTT
 AGTGTATTAGTCTGGCACCCCCCAGCCTGTGCTGTTCTGCTGACTTAGTCTATCTCAGGCATCTGACA
 30 CCCCCACAAGCTAACGCTTAAATATTGTTCTGCTGAGTGTGTTCTGAGCTTGGCCCTCCGCTTGTGTTCTGCTTCTG
 TTATTGCTGGTAAACCCAGCTTACCTGCTGGCTCCATGGCATCTAGCGACGTCGGGACCTCTGCTTATGATGC
 ACAGATGAAGATGTGGAGACTCACGAGGGGGCGTCATCTGGCCCGTGAGTGTCTGGAGCACCACGTGGCCAGCTT
 35 CTTAGCCAGTGTGACAGCAACGTCCTGGCTGGTTAGCCTGGAAAACCCAGGATGTCGGGCTCTGGGGAGTGTCTG
 CGCGGGTGTGAGTTGAAATCGCGCAAACCTGCGGTGCGGCCAGCTCTGACGGTGTGCTGGCTGGGGGGAGTGTCTG
 CTTCTCCCTCTGCTGGAACAGGACAAAGGATGAGGCTCCGAGCGTTGTCGCCAACAGGAGCATGACGTGAGCC
 ATGTGGATAATTAAAATTCTAGGCTGGCGCGGTGCTACGCCCTGTAATCCAGCACTTGGAGGCCAAGCGGG
 TGGATCACGAGGTCAAGGAGTGTGAGACCATCTGGCAACATGATGAAACCCATCTGACTAAACACAAAAATTAGC
 40 TGGCGTGGTGGCGGGTGTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTGAACTGGAGTTGGAA
 GTTGCAGTGTGAGCCACATTGACCAACTGCACTCCAGCTGGCAACACAGCAGAGACTCTGCTCAAAAAAAA
 AAAAAAAAATTCTAGTAGCCACATTAAAAAGTAAAAAGAAAAGGTGAAATTAAATGAAATAATAGATTCTGAA
 GCCCAGCATGTCACACCTCATCTTAAAGGTGTTATTGGTGGAGCATCACTCACAGGACATTGACATTGAGC
 TTTGCTGGGGATCCCGTGTGAGGTCCCGTGCCTGGCCATCGGGCTGGACCTGCTGGGCTTCCATGCCATGGCT
 GTTGTACCAAGATGGTCAGGTCCGGATGAGTCGCGCAGGCCCTGAGTGTGAGCTGGATGTGCACTGTCGGATGGTGCACG
 TCTGGGATGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGAGCTGGATGGTCTGGATGGTGCAGGTGAGGGTCA
 GAGGTCTCCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATG
 TGGTGTCTGGATGGTCAGGTCAAGGGTCTCCAGGCCCTCGTAAGCTGGAGGTATGGAGTCCGGATGATGCA
 GGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGTCTGGATGGTCAGGTCTGGAGGTGCCAGGTACCC
 AGGCCCTGCGGTGAGCTGGTGTGCGGTGTCTGGATGGTCAGGTCTGGAGGTGCCAGACGGTGCCAGACCATGC
 5 GGTGAGCTGGATATGCGGTGTCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGTCTGGATGGTCAGGTCT
 GTCCGGATGCTGCAAGTCGGTGTGAGGTCACTCAGGCCCTGCTGTGAGCTGGATGGTGTCTGGATGGTCAGGTCTGGCAG
 GGGGTGAAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGTCTGGATGGTCAGGTCTGGAGGTGCCAG
 10 GCCCTCGGTGAGCTGGATGGTCAGGTGTCCAGATGGTCAGGTCCGGGTGAGGTGCCAGACCCCTGCGGTGAGCTGGATG
 TGCGGTGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGTCTGGATGGTCAGGTCT
 GGTCCGGGTGAGGTGCCAGACCCCTGCTGTGAGCTGGATGGTCAGGTCTGGATGGTCAGGTCTGGAGGTGCCAG
 AGACCCCTGCTGTGAGCTGGATATGCGGTGTCGGATGGTCAGGTCACTCAGGCCCTGCGGTGAGCTGGATG
 GGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGAGGTATGGAGTCGGATGATGCAGGTCT
 15 GAGGTCTGGGGTGTGGTCGCCAGGCCCTCGGTGAGCTGGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTCAGGTCTGGGTGAGGTGCC
 CACCAAGGCCCTGCGGTGAGCTGGTTGTGCGGTGTCGGTGCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGCT
 TGAGATGTGCGGTGTCCCCGTGTCGGATGGTCAGGTCCAGGGTGAGGTGCTAGGCCCTGGTGGATGTGCGGTGT
 GTCCGGATGGTCAGGTCTGGGGTGAGGTGCCAGGCCCTGGTGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGC
 20 GGGTGAGGTGCCAGGCCCTGGTGGCTGGATGTGTTGTCCGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGC
 GCTGTGAGCTGGATGTGCGGTGTCTGGATGGTCAGGTGCCAGGCCCTGCGGTAGCTGGATATGCGGTGTCCGGATGGTCAGGTCT
 GTGTCGGGATGGTCAGGTCTGGGGTGAGGTGCCAGGCCCTGGTGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCT
 CCAGGGTGAGGTCACTCAGGCCCTGCGGTAGCTGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGC
 25 CCCGTGTGAGCTGGATGTGCTGTATCCGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGCTGGATG
 TGCTGTATCCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGCGGTAGCTGGATGGTCAGGTCTGGGTGAGGTGCC
 GGTCCGGGTGAGGTCACTCAGGCCCTGCGGTAGCTGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGCGGTGAGCTGG
 AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGCTGG
 ATGTGCTGTATCCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGCGGTAGCTGGATGGTCAGGTCTGGGTGAGGTGCC
 30 TGCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGGTGCCAGGCCCTGCGGTGGCTATGTGTTGTCTGGATGGTCAGGTCCGGGTGAGGT
 CGCCAGGCCCTGCGGTGAGCTGGATGTGTTGTCTGGATGGTCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTGAGC
 TGGATATGCGGTGTCCCCGTGTCGAATGGTCAGGTCCAGGGTGAGGTGCCAGGCCCTGGTGGGCTGGATGTGCGGT
 GTCCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGGTGGATGTGCGGTGAGGTGCCAGGCCCTGCGGTGAGGT
 GGGTGAGGTCACTCAGGCCCTGCGGTGATCTGGATGTGGCAATGTCCTTCTCGTTAAAG
 35 **Intron 3 (SEQ ID NO 6)**
 CTACTGTATCCCCACGCCAGGCCCTGCTCTCGAAGTCTGGAAACACCAAGGCCGCCCTCAGCATGCCCTGTCTCCACT
 TGCCCTGTGCTCCCTGGCTGTGCAAGCTCTGGCTGGAGGCCAGGGCCCGTCACAGGCCCTGGCCAAGTGGATTCTGTG
 CAAGGGCTGACTGCTGGAGCTCACGTTCTTACTTGTAAAATCAGGAGTTGTGCAAGTGGTCTCTAGGGTTGTA
 AAGCAGAAGGGATTTAAATTAGATGAAACACTACCACTAGCCTCCTGCCCTGGATGTGGGTCTGATTCTCTC
 40 TCTCTTTTTTTCTTTGAGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGAGTGGCATAATCTGGCTCACT

GCAACCTCCACCTCCTGGGTTAACGATTACCGCTCAGCCTCTAAAGTAGCTGGGATTACAGGCACCTGCCACAC
 GCCTGGCTAATTTGTACTTTAGGAGAGACGGGTTTACCATGTTGCCAGGCTGGCTCGAACTCATGACCTCAGG
 TGATCCACCCACCTTGGCTCCCAAAGTGTGGGTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT
 CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTTAGGTATGAGAGGATAAAATCCACCCACTGGGACTCACTG
 5 CAGGGAGCACCTGTGCAGGGAGCACCTGGGATAGGAGAGTTCACCATGAGCTAACTCTAGGTGGCTGATTTGAATG
 GCTGTGAGATTTGTCTGCAATGTTGGCTGTGAGAGTGTGAGATTGTGACAGACATTCAAGCTGGATTGATCAGTGAG
 GGACGGGAGGCCCTGGTCTGGGAGATGCCACCGTGGCTGAGCCAGGCTGGTATTAGCTTCTCCGTGCCCAGGCG
 TGACTGTGGAGGGCTTAGTCAGAAGATCAGGCTTCCCCAGCTCCCTGACACTCGAGTCCCTGGGGGCTTGTGAC
 10 ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
 GACCCGACGTGGTGCCTGGGCATTTCCTGATCTGGGGAGGGTCAGGGCTTCCCTGTTGGAAACAAGTTAACAC
 AATGACACCTTACTTAGACTTACACGTATTAAATGGTGTGCGACCCAACATGGTCATTGACCAGTATTTGGAAAGAAT
 TTAATTGGGGTGAACCGGAAGGAGCAGACAGACGTGGTCCCAGATGCTCTTGTCACTACTGGGACTGTTCTG
 CCTGGGGGCCCCCTGGAGGCCCCCTCCCTGGACAGGGTACCGTGCCTTCTACTCTGCTGGGCTGCGGCCCTGCGTC
 15 AGGGCACCAGCTCCGGAGCACCCGGCCCCAGTGTCCACGGAGTGCAGGCTGTCAAGCCACAGATGCCAGGTCCAGG
 GTGGCCGCTCCAGCCCCGTCCCCATGGTGGTTTGGGGAAAAGGCCAAGGGCAGGGTGTCAAGGAGACTGGGG
 CTCATGAGAGCTGATTCTGCTCTGGCTGAGCTGCCCTGAGCAGCCTCTCCGCCCTCCATCTGAAGGGATGTTG
 CTTTCTACCTGGGGTCTCTGGCTGGGCCAGCCTGGCTACCCAGTGGCTGTACCAAGGGACAGGCACTCTGTG
 AGGGCAGGGTACGGTGGCCCCAGATGAGCTGGGCTCTGGCTGAGCTGGGACAGTCACTGGGG
 20 GTTACGGCGCCGACTGGGCTCCCAGGGTGAAGTACAGGCTGCTGCGCTGAGCTGGGCTGCAAGTAGAGGGCTCTG
 AGGCCTCTGGCTGGCATGGTGGACGTGGCCCCGGCATGGCTTCAGCGTGTGCTGGTGGCTGAGCCCTCAC
 TGAGTCGGTGGGGCTTGCGCTTCCCTGAGCTTCCCCCTAGTCTGTGAGCAAGCCTCTGAGGGGCT
 CTATTGCAG

Intron 4 (SEQ ID NO 7)

25 GTGGCTGTCTGGTTAACCTCTTTAACAGAAGTGCCTTGAGCCCCACATTGGTATCAGCTTAGATGAAGGG
 CCCGGAGGAGGGCCACGGACACAGCCAGGGCATGGCACGGCCAACCCATTGCGCACAGTGAGGTGGCCAGG
 TGCCCGTGCCTCCAGAAAACAGCGTGGGGGTAGGGGGAGCTCTGGGGAGGGACAGGCTCTGAGGACCAAGAAG
 CAGCGGGCCAGGGCTGGATGAGCAGCACGGCCAGGTCTGGATCCGTGTCTGCTGGTGCAGCCTCCGTGCGCT
 TCCGCTACGGGCCCCGGGACCAGGCCACGACTGCCAGGAGCCCACGGGCTCTGAGGATCTGGACCTTGGCCACGG
 30 CTCCCTGACCCCCACCCCTGTGGCTGGCTGGTGGTACCCGTCATCTGAGGAGAGTGTGGGGAGGTGGACAGAG
 GTGTGGCATGAGGATCCCGTGTGCAACACACATGGGCCAGGAACCGTTCAACAGGGCTGAGGAAGCTGGGAGGG
 TTCTAGTCCCGGGCTGGTGGCTGGGACACTGGGGAGGGCTGCTCTCCCTGGTCCCTATGGTGGGGTGGCAG
 TTGGCCGGATCCACTTCTGACTGTCTCCCATGCTGCCCCGGCAG

Intron 5 (SEQ ID NO 8)

35 GTGGGTGCCGGGGACCCCCCTGAGCAGCCCTGCTGGACCTGGAGTGGCTGCCATGGCACCTCATGTTGGTGGAG
 GAGGTACTCTGGTGGGCCAGGGAGTGCAGGTGACCTGCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
 CTTCAGCCTTCTGCAAGCAGACATGGGGCCACTGTGCACCCCTGACTGCCCGGGCTCTATTCCCAAGGAGGGTCCACTG
 GATTCCAGTTCCGTAGAGAAGGAACCGCAACGGCTCAGCCACCCAGGGCCCTGGCTGACCCAGTCCGTGAGCCAG
 40 GGGTCTCCTGCTCTGAGGCTCAGAGAGGGACACAGCCGCCCTGCCCTGGGTCTGGAGTGGTGGGGTCAAGAGAGAG

AGTGGGGGACACCGCCAGGCCAGGCCCTGAGGGCAGAGGTGATGTCTGAGTTCTGCGTGGCCACTGTCAGTCTCCTCGC
CTCCACTCACACAG

5'-Bereich Intron 6 (SEQ ID NO 9)

5 GTAAGGTTACGTGTGATAGTCGTGTCAGGATGTGTCTCTGGGATATGAATGTGCTAGAATGCAAGTCGTGTCGTG
ATCGCTTCTGTGGTGGAGGTACTCCATGATTTACACATCTGTGATATGCGTGTGGCACGTGTGTCATGGTGTGTCCTGTGGTGCAT
GTATCTGTGGCGTGCAATTGTGGTGTGTGTCATGCGTGTGTCATGTCTGTGATGTGCCTATTTGTGGTGTGTCAT
TGTGTGTGTGTGACACGTGCATGTTCATGCGTGTGTCATGTCTGTGATGTGCCTATTTGTGGTGTGTCAT
GTGTCCTGTGACATATGCGTGTATGGCATGGGTGTGTGGCCCTTGGCCTTACTCCTCCCTCCAGGCATGGTC
10 GCACCATTGTCTCACGCTCTGGGTGCTGGTTGGGAGCTCCACATTCAAGGGCTCTACATTCTAGCATGGGTGCCCT
GTCTGTACAGGGCTGGGCTTGGAGACTGTAAGCCAGGTTGAGAGGAGAGTAGGGATGCTGGTGTACCTTCTGGA
CCCCCTGGCACCCCCAGGACCCCAGTCTGGCTATGCCGCTCCATGAGATATAGGAAGGCTGATTCAAGGCCCTGCTCCCC
GGGACACACTCTCCAGAGCGCCGGGGCTGGGCTGGCAGGGGTGAAAGGGCCCTGGCTGGGTTCCACCC
AGTGGTCATGAGCACGCTGGAGGGTAAGCCTCAAAGTCGCCAGGCCGGGTGCAAGGGTGAAGAAGTATCCCTGGA
15 GCTTCGGTCTGGGAGAGGCACATGTGAAACCCACAAGGACCTCTTCTGACTTCTGAGCT

3'-Bereich Intron 6 (SEQ ID NO 10)

TGTGGGATTGGTTTCATGTGTGGGATAGGTGGGATCTGTGGGATTGGTTTATGAGTGGGTAACACAGAGTTCAAG
GCGAGCTTCTTCTGTAGTGGGCTGCAAGGTGCTCAAACAGCTTATTGAGGAGACCATATCTTCTTGAACATATGGT
20 CGGGTTATAGTAAGTCAGGGGTGTTGGGCTCCCTGGCTCCCTGTTCTGTTCTTCCACTCTGGGTGCTGGT
CCTGCTGTGGTGTGGCCGGTGGCAGGGCTTCCAGGCCTCTGTTCTGTTCTGTTCTTCCACTCTGGGTGCTGGT
CCGTCTTGAATTCCCTGCGAGTTGGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
GTCTCGCTCTTTTGCCAGGCTGGAGTGGTTGGCGTGATCTGGCTACTGCAACCTGTGCTTCTGAGTTCAAGCA
ATTCTCTGGCTCAGCCTCCAAAGTAGCTGAAATTATAGCGCCACACCATGCTGACTAATTGGTAAATTAGT
25 AGACGAGGTTCTCCATGTTGGCCAGGCTGGCTCGAACCTCAGGTGATCTCCACCTCGGCTCCAAAGT
GCTGGGATGACAGGTGTGAACCGCCGCCGGCGAGACTCGCTCTGCGAGCTCCGTGAGATCTGAGCGATAGCTG
CCTGCAGCCCTGGTGTGACAACCTCGTTCTCCAGGTCTCGCTAGGGCTTCCATTCATGACTCTTCA
CAGAAGAGTTTACGTGTGCTGATTTCCCGCTGTTCTCGCTGTAATTGGTGTCTGTTATCGATGGCCTCTTCA
TTTCCCTTCTGCTTCTGTTATTGTTGTTCTCGCTCTGAAAGAAAAGTTGATTGGATGTTGAACCTTCTT
30 TCTAAACAAAGCATGAAAGTGTGGCTTCTAAAGCAGGGATCCCGAGGGCCCTGGCTGTGGAGTGGCACCGGTCT
GGGGCCTGTTAGGAACCCGGCGCACAGCGGGAGGCTAGGTGGGTTGTTGGGAGGCCAGCGTCCCGCTGAGCCCCCCCC
TCTCAGATCAGCAGTGGCATGGTGTGCTCAGAGGCACACACCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATTCT
GTGCTCTTATGGAACTTAATGCTGATGATCTGAGGTGAAACCTTGTCTCCAAACCATCCCTCCCCACTGCTG
TCCGTGAAAATCGTCTCCACGAAACAGTCCTGGTACCAATGGTGGGACCTGTGCTAAAGACCTGCTTCA
35 GCAGCCTCTGTCAGTGTGATATTGGCTTCTGTTGAGTCCAGAATAATTGAGTTCTGATGCTTCCGC
CGACCTCAGACCCATGGCTATTGTGGCGTGTGCTGCTCTGGTGGGAAGGGTGCAGGCCCATGTACCTTCC
GTTACTGCCTCCAGGTTGGTCTCAGGGTTGAATCGTACTCGATGTGGTTAGGCCACGGCCCTGCCAGCTCC
GGGGCTGGGAACATGCTGAAGCACAGACTACCGTGCCTGCTTGTGCTCACAGCTCGAGGCCCTGTGTC
TGTTAGTGTGTGTCACGTGCCGTCACTCTGTCTGGGACGCAGGGCTTAGCAGGTCCCGTAGTAAATGACAAGC
40 GTCCTGGGGAGTCTGAGAATAGGAGGTGGGGTGCCTCTCTCCCGTCTCAGACTCTTCTGCTTGTGCT

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGGTTGGGCACGGACTCCCAGCAGTGGGCTCCTCCCCTGGGCAATCACTGGGCTCATGACCC
GACAGACTGTTGGCCCTGGGGGAGTGGGGGATGAGCTGTGATGGGGCATGATGAGCTGTGCGCTTGGGAAATC
30 TGAGCTGGGCACTGCCAGGCTGCACAGCTGCTGCATTCAAGGCACCTGCTCACGTTGACTGCGCGCCCTCTCTCCAGTT
CCGAGTGCCTTGTTCATGTTGCTAAATGTTCTCTGCCAGTTGATCTGAGGCCAAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTGAGCCGTGCTGCCAGCTGGCCCTCACTGCTGGGCTGAGGCCAAAGGAAACG
TGTCCCCCTCTTAGGAGGACGGCCGTGTTGAGCCACGCCCGCTGAGCGGCCCTCACTGCTGGGCTGTCACAGT
GGCCCTGTGCCCTTGAGATGTTGCTGCCAGCTGGCCCTGTGGCTTTGAGATGCTGTTAGCACTTGCTCGG
35 TCTAGGGACAGTCGTGTCACCGCATGAGGCTCAGAGACCTCTGGGCAATTCTGCTCCAGGGGGGGGGGGGGGG
GTGGCCTGGGCTGCTGGGACCCAGACCCCTGTGCCCGAGCTGGGAGCAACTCTGGATCACATATGCCATCCGGGCA
CGGTGGGCTGCTGGGCTGAGCCAGCTGACCCACGGTGGCCAGAGGAGACCTCTGTCACACACTCTGCCAA
GCCCATGTGTCCTGCAGAGACTCGGCCCGCCAGCCACGATGGCCCTGCATTCCAGGCCAGCCCCGCACTTCATCACA
40 AACACTGACCCAAAAGGGACGGAGGTCTGGCCACGTGGCTCTGCTCAGCACCCACGGCTCACTCCCATGTG
TCTCCCGTCTGCTTCGAG

CGTTAGGGCTTGGGGAGATGGGCTGGTCAGCCTGAGGCCACATCTCCAGCAGGCCCTGACAGGTGGCTGGA
 CTGGCGCCTTCAGCCATTGCCATCCACTGATGGGCTACACCCAAGGACGCACACCTAAATATCGTGC
 AACCTAATGTGTTCAACTCAGCTGGTTATTGACAGCAGTTACTTTTTTTAATACTTTAAGTTAGGGTAC
 ATGTGCACGACGTGCAGGTTAGTTACATATGTATACTGTGCCATGTTGGTGTGCACCCATTAACTCATCATTACA
 5 TAGGTATATCCTAATGCTATCCCTCCCACTCCCCATCCCATGACAGGCCCTGGTGTGATGTTCCCCACCCCTG
 TGTCCAAGTGTCTCATTGTCAGTCCCACCTGTGAGTGAGAACATGTGGTTGGTTCTTCCTTGCAATAGTT
 GCTCAGAGTGTGGTTCCAGCTTCTGTCATGCCCACAAAGGACATGAACCTATCCTTTTATGACTGCATAGTATT
 CCGTGTGTATATGTGCCACATTCTTAACTCAGTCTATCATGATGGACATTGGTTGGTGCAGTCTTGCTACT
 GTGAATAGTGGCGCAATAAACATACGTGTCATGTTATAGCAGCATGTTATAATCCTTGGGTATATACCCAG
 10 TAATGGGATGGCTGGTCAAATGGTATTTCTAGTCTAGATCCTGAGGAATCACCACACTGTCTCCACAATGGTTGAA
 CTAGTTACACTCCCACCAACAGTGTAAAAGTGTCTGGTGGAGAGGATGTGGACAGCAGTTATTTTTATGAAAAA
 TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACAGCCATTCTCGAAC
 TCCGGGTTTCTGTGCATTTGAAACTCTAGCTCAATTATAGCATGTACAGTGGATCAAGGTTCTTCATTAA
 15 GGTCAAGTTCTAGATTGAAATAAGTTATGTAACAGAAAACAAAATTCTGTACACACAACTTGCTCTGGGATTGGA
 GGAAAGTGTCTCGAGCTGGCGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCATGGTCATGGGGCGCTGG
 GCTGGGCCTGAGGTCAACAGTCACAGTCACATGCCAGCTCTGTGGATAGGATCTGGTCTCGGATCATGCTGAGGAC
 ACAGCTGCCATGCTGGTAAAGGCACCACGTGGCTCAGAGGGGGCGAGGTTCCAGGCCAGCTTCTACCGTCTCAG
 TTATTTTCCCTAAGAGTCTGAGAAGTGGGCGCGCTGATGGCTTCTCGTCTCAGCTGGCACAGAATTGACAAA
 GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAAGTGTAGAGAGCTGCTGTTGGAAA
 20 GAAATTAAAGTTTCTTAACCGCTTGGAGATGTTACTTTATGGCTGTAAATTGTTGACATTCACTGCTGGGATTGGA
 TCGTAGACAGATACTACGTAAAAGTAAACCTGCTGTAAATTCCCTTATTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCCGTGCGTGTCTGTTGGGACCTCCACAGCCTGTTGAGGCCCTGTCCTGCCCTGG
 25 CACCGCAGCGTTGTCTCTGCCAAGTCCTCTCTGCGGCTGCTGGATCCGCAAGAGCAGAGGCCCTGGCGTGCACC
 CAGGCCCTGGGGCGCAGGGCACCTCGGGAGGGATGGTACCGTGAGGCCCTGGTCTCGCAGAGACGCCAGGTT
 ACACACGTGGTGAGTCAGCGGTGACCTGGCTCTGCTCTTGGAAAGTCAGAGTGGCGGCTCTGGGGCCAG
 TGAGACCCCAGGAGCTGTGCACAGGCCCTGCAGGCCAGGCCAGCTCCCAAGGTGCACCTGAGCCTGCCAG
 GAGCAGGAGCTGCTGAGTGAGCTGGCCACAGCGTCTGCGGTCAGTCTCGTGTGGGGTTGGATCGTGGGG
 30 AGAATTGGATTGCTGAGTGCTGCTTGAACCACGGAGATGGCTAGGAGTGGTTTCAAGAGTTGATTTTGAA
 CAAACTAAATCAGGCACAGGGACCTGCCCTCAGCACAGGGATTGCTCAATGTGGCTCCCTCAAGGGCGCCCAAG
 AGCCGGTGGCTTAAAGTCGATTGACGAGGGAGAGAAACCTTGAAAGCTGTAAGGGAACCTCAGAAAATG
 TGGCCGCCAGGGGTGGTTCAAGGTGCTTGTGGCTGTGTTTGAAACCCATTGGACCCCTCCAAGTCCACCC
 35 TCCAGGTCCACCCCTCCAGGCCCTGGCTGGGGTATGCCCTGGCTCTGTGCGCAGGCCAGCAGCAGGC
 TGTGCACATTAAATCCACTAAGATTCACTCGGGGGAGGCCAGGTCCAAGCAGACTGAGGCTCAGGAGTCCTGAGGCT
 GCTGAGGGGACAGAGCAGACGGGAACGCTGCTCTGTGTGCAAGTCTGAGGGTGTGGCCAGGGAGGTGGCTCAGA
 GTGTATGTTGGGCTCCACCGGGGGCAGAACCTCTGTCTGTGAGTCGGCAGGCCATGTAACAGGAAGGGTGGCCACAG
 GGAGCTGGGAATGCACCAAGGGAGCTGCCAGCTGGCCAGGGTCCACAGGCCAGAGGAAGGGCAGGGGGACGCC
 40 GGGGCCACAGCAGAGGCCAGGAAGGGAGGCCAGGCCAGAGCAGAGCTACCGGGCACAGGGGGCTCCCTG
 AGCTGGGTGAGCGAGGCTCATGACTCGGGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCCAGCTCACAG

CCCAGCCAGGTCCCGCCCTGAGCAGGAACCTCAGAACCCCTCCCTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
 AGGGAAAACAGGCAAAGTCGTTGAGAAACGTCTAAAGAAGGTGGATGGTGGCAATTCTGTCTCAGATTTAGTCT
 5 GCCCCGGACACAGATGAGTCTATAACGGGATTGTGGTGGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
 TGGGGCTGCACCTCCCATCTGAGTCCTGGCTGCCAGGGCAGGTCTTGCATGCTCACCTACCTGTCTGCC
 GGGAGACAGGGAAAGCACCCCGAAGTCTGGAGCAGGGCTGGTCCAGGGCTCTCAGAGCTCTGCCAGGCCAGCACCC
 GCTCCAAATCACCACCTCTGGGTTTCCAAAGCATTAAACAAGGTGTAGGTACCTCCTGGGTGACGGCCCCGCA
 TCCTGGGCTGACATTGCCCTCTGCCCTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGCTGGGAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCT
 TCCGTGGGAGGCAGTCGAACTCCAAAGGTCAAGGCCACAGGGTCCCCCTGCTCCATCTGGGCTGAGCAGA
 AATGCATCTTCTGTGGGAGTGAGGGTGTCTCACAAACGGGAGCAGTTCTGTGCTATTGGTAAAGGAATGGTGCAC
 CAGACCTGGGTGCACTGAGGTGTCTCAGAAAGCAGTCTGGATCGAACCCRAAGACGCCGGGCCCTGCTGGCGTGA
 15 CTCTCAAACCCGAACACAGGGCCCTGCTGGCATGAGTCCTCTGAAACCCGAGACCTGGGCCCTGCTGGCGTGA
 CTCTCCGAACCCAGAGACTTCAGGGCCCTTGGCGTGAATCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
 AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGACAGGGCATGGTGTGGGGGGGTCTACAA
 AATTCTGGGTCTTCTGGGAGAGCCAGCTAACGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA
 CACAGATGCAGAAATCTGTGTTCTTATGAATAAAAAGTATCAACATTCCAGGCAAGGTGGCTCACACCT
 20 ATAATCCCAGCACTTGGGAGGCCAGGGTGGATCACTGAGGCCAGGAGTTGAGGCCAACCTAACCAACATAGTG
 AAATTCCATTCTACTTAAAAAAATACAAAAAATTAGCCTGGCCTGGGCACACGCCCTGAGTCCCCGCTATGCCGGAGG
 TGAGGCAGGAGAATCATTAACCCAGGAGGAGGTTGAGCTGAGCCAGATCACCCACTGCACTCCAGCCTGGCA
 ACAGAGTGAGACTCATTTAAAAAAAGTATCAGCATTCCAAACCATAGTGGACAGGTGTTTTTATT
 25 TGTCTTCGATAATATTACTGGCTGTCTAGAGGCCAGTGGGTCTCTGAAAGGCACACCTCATGG
 GAAGAGAAATAAGTGGTAATGGTTAAACAGAGGTTAACTGGGTCTGTGAGTTAACAGTCCAGTC
 TGGACTTGCCTCTTCCAGAATGCTCCCTGGGTTGGCTCATGGGGAGCAGCAGGTGTTGACACCCCTCGTGA
 GAGCAGCAGGTGCAAGCCTCATGATGGGGAGTGGCAGGTGAGACACCCCTGTGCTGGACACCCCTCGTGA
 30 TTGAGCTCCCTCCCCAACAGGATGCCGCTCTGTGCTGATGATTCCACATTCTGGCTCCAGCACCCTCTGCC
 CTGGCCTCCACTGGCTTGTGCTGATGATTCCACATTCTGGCTCCAGCACCCTCTGCCCTCCAGGACCTCT
 GCAGTGCTGGCCATACCAGTCAGCTGTGAATGTCCTACTGCTTATTGCTCCCATGAAATGTATTAGGACAGGC
 ACCCCCTGGTTCCAGCCTGGCAGCAGTCAGTGAAATGTTATTGAGAGAACAGAACAAACAAATCAGGAA
 35 GTTCTCTAAACACATTGCAAAGCCACAGGGCTAGTGCAGGATGGTGGCATCAGGTGATGGTGGTCA
 CCAGAAATATTCTGTGCTCCCAAGGCCACTGGTCAAGGTGCTTGAGGGTGGCTCTAAAGCTCAGCAGTGGAG
 GCAGTGGTTCGCCATCTCAGGGTGAACTCACATCCTCTGTGCTGAAGTATACAGCAGAGGCTTGAAGGG
 40 GAAGAAAACAGGCAAATGATTAAGAAAAGTGAAGAAAAGTGAAGATGGAAATTCTGTCCAGATTAGT
 TCCCAACACAGTCAGATGGTAGAATGTGGTCAAGACTGATGGACAGAACATAGAACAAACGGAAAGCC
 CAGAACCTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGGTGTTGAGTGTG
 GACTGGAAAGCAAATAAGTGTGCTTACAGCATATACAGAGCAGATTCTAGGTAGAAGAGGAGACACATG
 ACCAGAACAGAAATAAAACAAAGACTCAAAGGAAGGGAGGTGAACGTTCCCTGGTTGGTGTGGGG
 AGGGAGGCCGATGAAACCACTGAGGCAACGGCATTGCTTCACTGAGAACACTCAGTTGCC
 ATGGCCATTCCCTGGAGCGTTGTGACGTGATTATTAAAGGCGCCCTGTGAGGTCTGCACATT
 CACTCCTCACTT

GTTCTCTAACCAACCACCTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTGGTCACCCAGCTGGCAAAGGGC
 ATGCATGATTGCGACCTGGCTCCTGCTCCGGGCCCTGCTCTGCCAGGACCCACACAAGTCAGACCCATAGGCTC
 AGGGTGAGCCGGAGCCAAAGGTGTTGGGATGGCTGTGAAAGAAGAAATGGACGTCGTGACACTTGGGAAGGTC
 CTACCCAGCAGCGTCAAAGAAATGCATGAAACTGACAGCGAGACCCATCCCTCAAAGAACGACGTGAAACTGATGGC
 5 GAGACCTGTCCCCATCCCTCATGCTGGCTCTTCTGGCTTGCCAAAGAGGCCAGCATCAGTTGAGGCAAGCTGGAAAG
 ACTTTCTGGAAAGCAGCTTGTGATGGAAGTCTCACAATGCTGTCTCCAGTAATTCACTTCTGAAGTGA
 CCAGACATTATCACGGGTCTTATTACATTCCAGTGTCCAGGCAGGGGACTGCCACAGCAAGTCACGAACCTGCC
 CAAATAACAGGGCTAAGGAGATTTATGCATCACAAAATGCTCTGCCATTAAACATTTCAAAGAATTGGAAAGAAT
 10 GTTTAATGGCACAAAAGTTTCAATGTCAGTGTCAAAGCTGGATGTAAGGAAACACACCCAGGAGCCTGCCG
 TGAATGTCATGTGTTCATCTTGGACATGGACATACATGGCAGTGAGTGGTGGTGGAGGCCCTGGAGGACATCGGTGG
 GATGCCCTCCATCCCTGGCCCTCTGGAGACACCATGTGTGCCACGTGCACTCACTGGAGCCCTGTTAGCTGGTGCACCTG
 GCTCTTCATCCCTGAGATTCAAACACAGTGAGATTCCCACGCCAACACTAGTGTCTCCACAAAACCTGAGTCAC
 ACCTGTGTTCACTCGAGGGACGCCGGAGCCAGGGCTCACAGTTATTATGTGTTTGGCTGAGTTATGTGAGATC
 15 TCATCAGGGCAGATGATGAGTGACAAACACGCCGTGCGAGGTTGGATAACACTCACACTAGCCAGGTCTGGT
 GAGTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTTGGAGTCCATGGAGTGAGCACCCAGGCCCTGGCTGAGC
 GCATGCCCAAGGCAAGGAAGCGGGAGGAAGGCAGGAGGCTTTGGAGCAAGCTTGCAAGGAGGGCTGGGTGT
 GGGCAGGCACCTGTGTCACATTCCCCCTGTGTCAG

Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGCGCAAGTATGTGTGTTGTCGCGCGCT
 GCCTGCAAGGCTGATGGTACTGGCTGCACGTAAGAGTGCACATGTACGCATATACACGTGAGCACATACATGTGTCAT
 GTGTGTCATGAAGGCATGGCAGTGTCAGGTGTGCAAGGGCACAAGTGTGTCACATGCAATGCAACACCTGACA
 TGCACTGTGTTCGTGCACAGTCGTGTGGCATTACGTGAGGTGCATGCGTGTGGTGTGCAGTGTGAGTAGCATGT
 GCACATAACATGTATTGAGGGTCTCGTGTGTTACCCCGCTAGGTCTCAGCACCAGTGCACACTCCTACAGGATGAGAC
 25 GGGGTCCTCAGGCCCTGGTGGCTGAGGCTCTGAAGGCTGCAAGCTGCAGCCCTGAGGGCATTGTCCCATCTGGCATTCCGCTCACT
 CCCTCTCTGTGGCTCTGTGTCACCTCCCTCTGTGGCATTACATCCACTCCACTCCCTCTGTGGC
 ATCCCGCTCCACTCCCCCTCTGTGGCATTCTGCGTCCACCTCCCTCTGTGGCATTTCGCTCCACTCCCTCTCC
 GTTCTCTGTCTGGCCAGCCTGGGGCAGGAGACACAGAGTCTGACTGCCAGGGTGGTTCGAGCTG
 30 CCGGGTGGAGGCCAGGCCGATTCACTGGGAAGAGGGATAGTTCTGTCAAATGTTCTTGTCCATCTG
 ATGGATGATAAAGCAAAAGTAAAACCTAAATCCAGAGGTTCTACCGTTCTACCTTGTGCACTCTAG

Intron 13 (SEQ ID NO 17)

GTGAGCCGCCACCAAGGGTGCAGGCCAGCCCTCCAGGGACCCCTCCGGCTCTGTCACCTCTGACCCGGGCTTCACCT
 35 TGGAACCTGGTTAGGGCAAGGAATGTCTACGTTTCAGTGTGCTGCTGCCGTGACAGTTCTGTTGCGCTG
 GCTCTGTGCAAAGCACCTGTTCTCCATCTGGTAGTGTAGGGAGCCGTGTGGCCCTAGGTGTCCCCACTGTGCTGT
 GCACTGGCGTGGACGTCATGGAGGCATCCAGGGCAGAGGGCATGGGTAAGAGATGTTATGGGGAGTCTT
 CAGAGGAGGCTGGAAAGGTGCTGAACAGTAGATGGAGATCAGATGCCGGAGGATTGGGTCTCAGCAAAGAGGGCC
 GAGGTGGGTGAGGTGAGGGTCGCTGGCCACCCCGGGAAAGGTGAGCAGAGCTGTGGCTCCACAGCCGGCCA
 40 GCACCTGTGCTCTGGCATGGCTGTGTCCTGGAACGTTCCCTGTGCTGGTCAAGGGGTGCCAGAATCG
 ACAACTTATCACAGAGGGAAAGGGCAATCTGTGGAGGCCACAGGGCAGCTCTGCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTGGGGCTGTACCAAAGGCAGTCGGCACACAGGCCGGCCTCACCTAACAGGCCCTCCAGGCCACTG
 GGAGCTGAATGCCAGGAGGCCAAGCCCTGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTTACCCAGGCC
 AGGCTGCGCAATTACCGTCACACTTGATGTGAAATGAGGTGCGTCTATCGTGGAAACCCAGCAAGGGCTCACGGGA
 GAGTTTCCATTACAAGGTGTAACATGAAAATGGTTTAACCCGAGTGCTTGCCTCATGCTCTGGCAGGGAGGGC
 5 AGAGCCACAGTCATGTTACCGCTTGTGACCAGCTCAGGGCTGTCAGTCCAGGGCTGCGTCTCAGTCCAGGGTGC
 GGCTCAGACGCCCTCCTCTGCTTCTCTGCTCAAATCTCCCTGTTGATCTCCCTGACGCGTGCCTGG
 CCCTCGTGCAAGCTGCTTGACTCTTCCGAAACCCCTGGGTGCTGGATAACAGGTGCACTGAGGACTGGAGGTG
 CTGACACTGTGTTGACCCCAGGGTCCAGCTGGCGTGTGGGCTCTGGGCATGAGGTGAGAGGAGTTTCC
 10 CAGGTGAAAACCTCTGGAAACTCCAGGGCCTGACCTGGGCTAGAGGGCTGGCTCAGGGCAGGGCGCTGAGTTCC
 ATTTCCCCACCAAGGGTCTCTAGCTCGAGGAGCTCCGAGGGCTAGAGGGCTGGGCTCAGGGCAGGGCGCTGAGTTCC
 CCATGTGGGACCCCTGGGTAGTCGCTGATTGGTAGCCCTGAGGAGGCGAGATGCGATGGGACCGGGCGTTCCA
 AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTCCCTGAGGGAGTGGGAGAACGGAGGCTGGCCCG
 ATTTACGGCAGCCAGGCTGAGTGGCGAGGGCTGTTGGGCTCAGGGCTGGGCTGGGCTGAGTTCC
 15 TGGGCTCGGCTTCTGGCCGTGCTGGCGCCTCCACAGGGCTGGGCTGGGACGCCGACCTCTAGCAGGTGGC
 CCGGGACCTTGGCTTATTTGGTAAACATCTGGGCTGGCTTGGCTGCTAAATGGGAAAGACATCC
 CACCTCAGCAGAGTTACTGAGAGGCTGAAACGGGGTCTGGCTTGTGACTGGTGTGATCTCAGGTCA
 CAGGAAGTCAGTGAGACCAGGTACATGGGGCTCAGGAGTGGGAGATGAGGTACACGGGGCTCAGGAGTGG
 20 GAGGCCAGGTACATGGGGCTCAGGACTGGTGTGAGATGAGGTACACGGGGCTCAGGAGGCTGAGACCAGGTAC
 ACGGGGCTCTGATCACACGCACATATGAGCACATGTGACATGTGCTGTTCATGGTAGCCAGGTCTG
 CACACCTGC
 CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGGGCTCACACCTGTAGTCCCAGCA
 CTTTGGGAGGGCGAGGCGAGAGGATCCCTGAGCCAGGAGTAAAGACCCAGCCTGAGCAACATAGTAGAACCC
 25 TATGAAAATAAAACAAAATTAGCTGAACATGGTGTGCGCTGTAGTTCAATACTTGGAGGCTGAAAGTGGAG
 GATCACTTGAGCCCAGGAGGTGGAAGCTGAGCTGAGCTGAGATTGACACTGACTGAGCCTGGG
 GACAGAGTGA
 GCCCATCTCAACAACAAGAAAGACTGACAATGCACTGGAAAGAAACATTAGTAGGAACCTAACCTACACA
 CAGAAGCCAAGTGGTGTCTGGTGTCACTGAGATGAGATGATGGGTCTCACACCATACCCAGACCCAGGGTTATG
 30 CACCAAGGGGGGGTGGCTCAGAAGGGATGCGCAGGAGCTGATACAGTGA
 CAGTGA
 GATTCACTGATAAGTACCTGCTGGTACACAAGGAACATGGATAAAACTGGAAACCTTAGAGGCCTCC
 CGAACAGGGCT
 AATCAGAAGCCAGCATGGGGCTGGCATCCAGGATGGAGCTGCTCAGCCTCCACATGCGTGT
 CATACAGATGGTGCA
 CAGAAACGCAGTGACCTGTGACACACAGACAGCAGCTACTGCACACACAAGCACACACAGACATGC
 ATGCATGC
 ATCCGTGTGTGACCTGTGACACACAGACAGCAGCTACTGCACACACAAGCACACAGACATGC
 ATGCATGC
 GCCCACACCCACGAGCACCGTGTGATTAGGAGGCCTTCTGACGCTGTCCGCATCCTCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGCAGGTGCTGGCTCAGTGGCAGCAGTGCCTGCTGGTGTAGTGTGTCAGGAGACTGAGTGAATCTGG
 CTTAGGAAGTTCTTACCCCTTCTGATCAGGAAGTGGTTAACCCAAACCTGTCAGGCTCGTCTGCCCGCCCTCTCGT
 GGGGTGAGCAGAGCACCTGATGGAAGGGCAGGGAGCTGCTGGAGCTGCCATCTTCCCACCTGCTCTGG
 GCGCTGGGGGGCTGGTCTCTCTGTTGGCCCATGGTGGATTGGGGGCTGGCTCTCTGTTGGCCCTGTGG
 GATTGGGCTGCTCCCGTCCATGGCACTTAGGGCCCTTGCAAACCCAGGCCAGGGCTTAGGAGGAGGCCAGGCCAG
 40 GCTACCCACCCCTCTCAGGAGCAGAGGCCAGTACCCACGAGGCCGCTCTGCTTCCCAGTCACCG

TCCTCTGCCCTGGACACTTGTCCAGCATCAGGGAGGTTCTGATCCGTCTGAAATTCAAGCCATGCGAACCTGCGGT
 CCTGAGCTTAACAGCTTCTACTTCTGTTCTTCTGTGTGAGGAAATTCAACCTGGAGAAGCCAGAAAACATTCTG
 TCGTGACTCCTGCGGTGCTGGTCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGAGCTTCCG
 GTGTCTCCTGGGAGGGGAGCTGGCTGGCTGTGACTCCTCAGCCTCTGTTTCCCCAG

5

Intron 15 (SEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTCGGGCCCCACCTGCCAGGGTCATCCTGAACGCCCTGTGTGGGCAGCAGCCTC
 AGATGCTGCTGAAGTGCAGACGCCCGGGCCTGACCCCTGGGGCCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG
 CTGGTGTCCCAGGCCACGGCCTGGCAGGGTCCCAACTCTTGAACCCCTGCTTCCATCTCAGGGCGATGGCTCC
 10 CCACGCTTGGGAGCCTCTGACCCCTGACCTGTGTCTCTCACGCCTTCCCTGGCTGCTGCCCTGAGCTCTGGGT
 CCTGAGCAAGTCTCTCCCCGCCCCGCTCAGCGTCACTGGCTGCTGTCTGCCCTGGGAGGGGTGCTG
 TCCCTTCACTGAGGTTCCACAGGCCAGGGCACAGGGTGCAGGCCCTGCCCTGCCACACGTCTAGGAGGG
 TTGGAGGATGCCACCTCTGCCCTTCTGAACGGAGTGTGATTGGCCCCGAG

15

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTGAATCTAAATGTCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGTGGT
 GGTCAGTGCAGGGCCATGGCTGGCTGTGATTTACGGAAGTCTATGAGTGAATGGGTTGTGGTCAGTGCAGGGCCATG
 GCCTGGCTGGGCTGGAGGTTCTGATGCTGTGAGGCAGGGAGGGAAAGGAGGGTAGGGATAGACAGTGGGAGCCCCCA
 CCCTGGAAGACATAACAGTAAGTCCAGGCCAGGGCAGCAGGGATGCTGGGGCCAGCTGGGGCGGGATGATG
 20 GAGGGCCTGGGAGGGTGGCAGGGATGATGGGGCCCCAGCTGGGTGGCAGGGTGTGATGGGGGGCTGGCTGGGTG
 CGGGGAAGATGGGAAGCCTGGCTGGCCCCCTCCTCCCTGCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
 GGTGACATCCTCTGGGCCATCAGCTTCACTGGGGCAGGGCATGACACCATCCTGTATAAAATCCAGGATT
 CCTCCTCTGAACGCCCAACTCAGGTGAAAGTCACATTCCGCTCTGCCATTCTTAAGAGTAGACCAAGGATTCTG
 ATCTCTGAAGGGTGGTAGGGTGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGCTGGTGTGCTCTC
 25 ATCCCTTATCATCTCCAGTCTCATCTCATCTCTTATCATCTCCAGTCTCATCTGCTTCCCTTATCTCCAGT
 CTCATCTGTCACTCTCTTACCATCTCCAGTCTCATCTCTTATCTCTTATCTCTAGTCTCATCCAGACTTACCTCCA
 GGGGGGGTGCAGGCTCGCAGTGGAGCTGGACATACGTCCTTCCTCAGCAGAAGGAACCTGGAAAGGATTGCAAGAACAG
 GAGGGCGGCTCAGAGGGACGCAGTCTGGGTGAAGAACAGCCCTCTCAGAAGTTGGCTGGCCACACGAAACCG
 AGGGCCCTGGTGAAGTGGCTCCAGAGCCTCCAGCAGGTCCTGGTGGGGCTTATGGTATGGCCGGTCTACTGAGT
 30 CACCTGGACAGGGCTCTGGTTGAGTGCAGGCCAGCTGGCTGGGGCTTATGGCCACTGGATATG
 GCGTCATTATTGCTGCTGCTCAGAGAAATGTCTGAGTGCAGCGAGCCTAATGTGTATGGGGCCAAAGTCCACAGACTG
 TGTCGTAATGCACTCTGGTGCCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGCTTGGCAGCCGGCTGGG
 GCGCCTTGCCTGCAAACGGAGGGAGCGGCCGGCCGTGGGGAGACCTCAAGTGAAGGAGGGTGGACAGAAC
 AGGGCGGGGACTTCCAGGAGCAGAGGCCGCTGTCAGGCACACCTGGTTGAATCACAGACCAACAGGTCAGGCCATT
 35 GTTCAGCTATCCATCTTACAAGCTCAGATTCTGTTCTCCGGTGTGTTTGTGAAATTTCAGGATTACT
 TATATTTTTGCTAAAGTATTAGACCTTAAAAAGGTATTGCTTGTATGGCTTAACACTAAGCACCTACTTTAT
 TTGTCTGTTTATTTATTATTATTATTAGAGATGGTGTCTACTCTGTCACCCAGGTTGTAGTGCAGTGGCAG
 AGTCATGGCTCGCTGTAGCCGAAACCCCCAGGCTCAAGTGTATGCTCCGGCCTCAGCTCCAGAGTGTGGATTACAG
 GTGTGAGGCCACTGCCCTGGCTGGCACTTTAAAACCACTATGTAAGGTCAAGGTCCAGGTCAGGTTGAGACCAG
 40 CAGTAGTTGGAAGCCGAGGCAGAAGGATTGCTGAGGCCAGGAGTTGAGACCAGCATGGTAACATAGGGAGACCC

ATCTCTACAAAAAATGCAAAAGTTATCCGGCGTGGGTCCAGCATCTGAGTCCCAGCTGCTCGGAGGCTGAGTGGG
AGGATCGCTTGAGCCGGAGGTATGGCTGAGTGTGATTGACCATCGCACTCCAGCCTGGCAACAGACTGA
GACCTGTCTCAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGAAGAAGGAAAGAAGAAGAAGAAGAAGA
GAAGAAGGAAGAAGAAGGAGAAGGAGGCTGCTAGGTGCTAGGTAGACTGCAAATCTCAGAGCAAATGAAAATAACA
5 AAGTTTAAAGGGAAAGAAAAACCCAGCTCTTGACTTCTAGGCCTGAACCTCATCTCAAGCAGCTTCTTCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAACGAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTCTGCCTCAGCCCCACGCTCCTGCCGGCTCTGCACCTGCTGAACCGC
GATGTTGGTCCCAGGTGCCCACCTGGAAAGGATGCTGTGCAGGGGCTGCCAAACTTGGTGGTTCAAGAGCCCG
GCCACTTGTCGGCAGGCACAATTACAGCCCCCTCCCCAAAGATGCCACGTCTCTCTGGAACCTGTGAATGTGTCACCCG
10 CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTAAGGTATCTGGATTATCTGG
TGGGCCTGATATGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTGAGATGGAGGAGGGGCTCCAGCCAAGGAATGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCCGGTCTGAGGGCACACGGCCCTGCCACGCTGATTCAGGCCAGTGGACCTGTTCAAGCTTCAGCTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTGTGTTCAAGCCACTAAGCTGCAGTGATTGTCACAGCAGCAAATGGAATAG
15 CAGTACAGGGAAATGAATAACAGGGACAGTTCTCAGAGTGACTCTCAGCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-5 charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative 15 Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

15

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 20 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des 25 offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

25

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von 30 Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalierte Splicesstelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, lässt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplittet wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpacket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur 10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv markierter Primer (5'GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵ cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL), 2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl 15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden 2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150µl 20 Phenol/Cloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-Aacetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stoplösung gelöst. Nach 5 min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel 25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom 30 ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA_nTTGT) repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_nNa/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzutragen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den 10 Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem 25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 30 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGAAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-
10 CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

20 In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. 25 bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

30 Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpacket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschiedene lange hTC-Promotorsequenzabschnitte generiert und 5' vor 10 das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 15 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACCGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTCC-3', Position -4093 bis -4070) 25 wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaars PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination 30 PK-3A und PK-5D (5'-

GGAGATCTGATCTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGCTGGATTCCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes 5 (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. 25 CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit 10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszuplattieren.

15 Für die Transfektion wurden pro Well jeweils 2x 10³ -HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV 20 β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serumfreies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl 25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄ Mg(OH)₂·5H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysat und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na₂CO₃ gestoppt und die Absorption bei 420 nm bestimmt.

5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotor-sequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK 10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im 15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

5 Allsopp, R. C., Vazire, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.

10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Whiley-Interscience, New York.

15 Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.

20 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.

25 Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.

30 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.

35 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthetase promoter. Mol. Cell. Biol. 14:4894.

40 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.

45 Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.

5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.

10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.

15 Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.

15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.

20 Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.

20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.

25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.

30 Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.

30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.

35 Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.

35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 **Shay, J. W. (1997).** Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

10 **Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994).** Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

15 **Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995).** Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Jurnal* 311, 549-554.

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.

5

2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.

10

3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.

15

4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.

20

5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.

6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.

25

7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.

8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

30

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:

5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,

10 B. Messung des Substanzeffektes auf die Reportergenexpression.

10 10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.

15 11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.

20 12. Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

25 A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält mit Körperflüssigkeiten oder zellulären Proben,

 B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

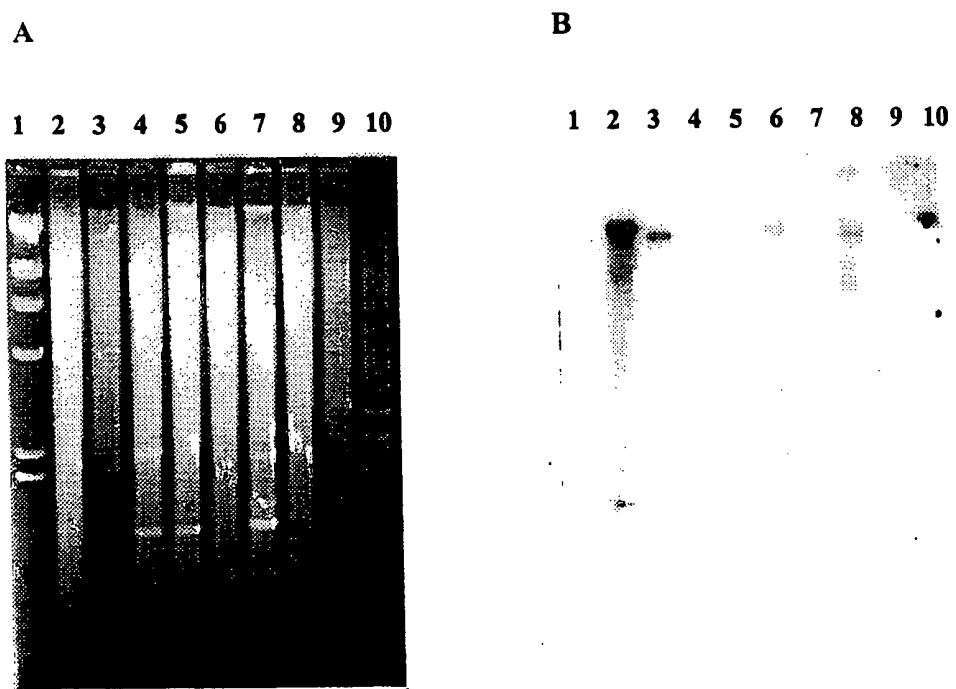
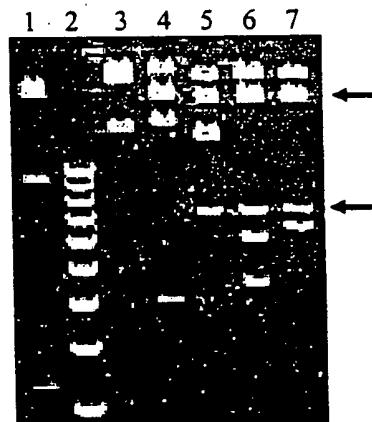
Fig. 1**ERSATZBLATT (REGEL 26)**

Fig. 2



ERSATZBLATT (REGEL 26)

Fig. 3

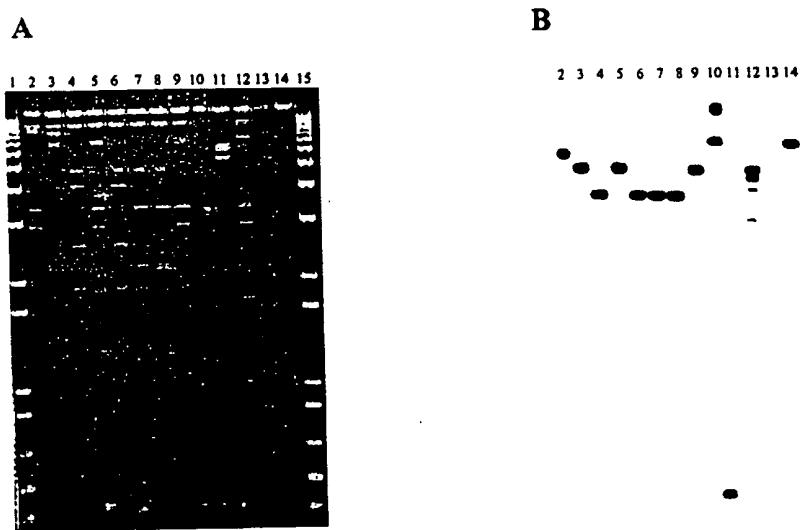
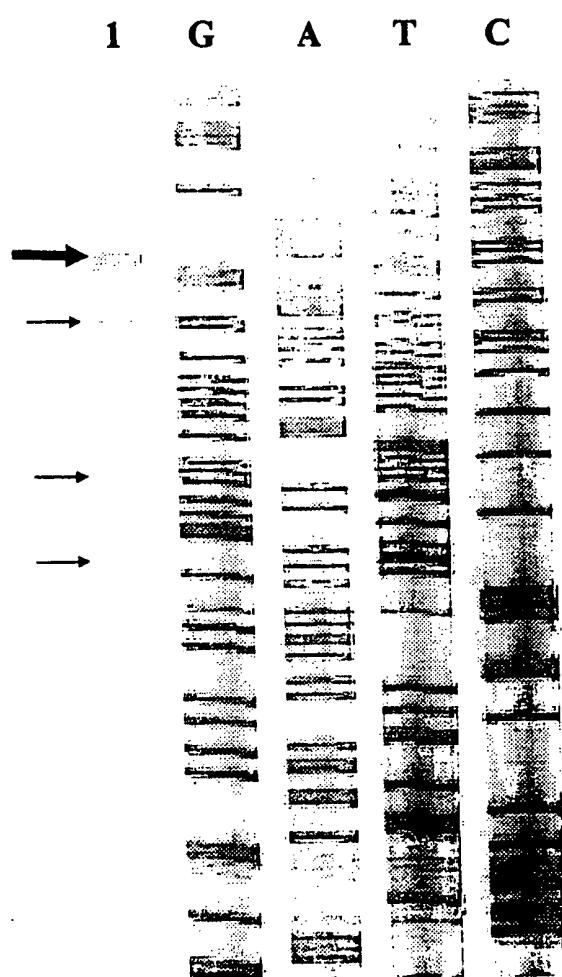


Fig. 4

Fig. 4 (Fortsetzung)

CACAGCCTAG GCCGATTCGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCCT GGGAGCGCGA 4760
GGGGCGCGCG GGCAGGGAAAG CGCGGCCAG ACCCCCAGG CCGCCCGGAG CAGCTGCCT GTCGGGGCCA 4830
GCCCGGGCTC CCAGTGGATT CGCGGGCACCA GACGCCAGG ACCGGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCGGGGCCCT CCCAGCCCCC CCCCCCTCTT TCCGGGGCCCC 5040
CCCCCTCTCC TCGCGCGCG AGTTTCAGGC AGCGCTGGGT CCTGCTGGCG ACGTGGGAAG CCCTGGGGCCC 5110
GGCCACCCCCC GGGATG 5126

Fig. 5



ERSATZBLATT (REGEL 26)

Fig. 6

GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG CGATGCCGCG 70
 CGCTCCCCGC TGCGGAGCG TGCGCTCCCT GCTGCGCAGC CACTACCCCG AGGTGCTGCG GCTGCCACAG 140
 TTCTGCGGC GCCTGGGGCC CCAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCCGGGGCGCT TTCCGCGCGC 210
 TGGTGGCCCA GTGCGCTGGT TGCGTGCCT GGGACGCAGC GCGCCCCCCC GCGCCCCCGCT CCTTCGGCCA 280
 GGTGTCTGC CTGAGGGAGC TGGTGGCCCCG AGTGCTGCG AGCGCTGCG AGCGCGGGCG GAAGAACGTG 350
 CTGGGCTTCG GCTTCGCGCT GCTGGACGGG GCGCCGGGGG GCGCCCCCGA GGCGTTTACCC ACCAGCGTGC 420
 GCAGCTACCT GCCCAACACG GTGACCGACG CACTGGGGG GAGCGGGGGG TGGGGCTGCG TGCTGGCCCG 490
 CGTGGGCGAC GACGTGCTGG TTACCTCTG GGCACGCTGC CGCTCTTGTG TGCTGGTGC TCCCAGCTGC 560
 GCCTACCAAGG TGTGGCGGCC CGCGCTGTAC CAGCTGGCG CGCCACACTCA GGCGCCCCCCC CGGCCACACG 630
 CTAGTGGACC CGGAAGGCGT CTGGGATGCG AACGGGGCTG GAACCATAGC GTCAAGGGAGG CGGGGGTCCC 700
 CCTGGGCTG CGAGCCCCGG GTGCGAGGAG GCGCGGGGGC AGTGCAGGC GAAGTGTGCC TTGCCCCAAG 770
 AGGCCAGGC GTGGCGCTGC CCTGTGACCG GAGCGGGAGC CGCTGGGGCA GGGGGCTGCG GCGCACCCGG 840
 GCAGGACGGG TGAGGGAGT GACCGTGGTT TCTGTGTGGT GTCACCTGCG AGACCCGGG AAGAACGCCAC 910
 CTCTTGGGAG GTGCGCTGCT CGCGCACCGG CCACTCCCGA CCACTCGTGC CGCCGCAAGCA CGACGGGGC 980
 CCCCCATCCA CATCGGGGCC ACCACGTCCTC TGGGACACCC CTGTCCCCC GGTGTACGCC GAGACCAAGC 1050
 ACTTCTCTA CTCCCTAGGC GACAAGGAGC AGTGCAGGC CCTCTTCTA CTCAGCTCTC TGAGGCCAG 1120
 CCTGACTGGC GCTCGGAGC TCGTGGAGAC CACTTTCTG GTTGGTACGG CCGTGGATGCC AGGGAATCTCC 1190
 CGCAAGTGC CGCGCGCTGC CGACGCTAC TGGCAAATGC GGCCCCCTGT TCTGGAGCTG CTTGGGAACC 1260
 ACCGCGAGTG CGCCCTACGGG GTGCTCTCA AGACGACTG CGCGCTGGCA GCTGGGGTCA CCCAGCAGC 1330
 CGGTGTCTGT CGCCGGGAGA AGCCCCAGGG CTCTGTGGG GCGCCCCGAGG AGGAGGACAC AGACCCCGT 1400
 CGCCTGGTGC AGTGTCTCCG CGACGACAGC AGCCCTGGC AGGTGTACGG CTTCTGTGGG GCGTGGCTGC 1470
 GCGGGCTGGT GCGCCCGAGC CTCTGGGGCT CGAGGACACG CGAAGCCCGG TTCTCTAGGA ACACCAAGAA 1540
 GTTCATCTCC CTGGGGAGA ATGCCAAGCT CTGGCTGCGAG GAGCTGACGT GGAAGATGAG CGTCCGGGAC 1610
 TGGCTTGGC TGCGCAGGAG CGCGGGGGT GTCTGTGTG CGGGGGCGAGA CGACGGCTCTG CGTGAGGAGA 1680
 TCCCTGGCAA GTTCTGTCAC TGGCTGTGAG GTGTGTACGT CGTCAAGCTG CTCAGGTCTT TCTTTTATGT 1750
 CACGGAGACC ACGTTTCAA AGRACAGGGT CTTTTTCTAC CGGAAGATG TGAGGAGCAGA GTTCCRAAGC 1820
 ATTGGAAATCA GACAGCACTT GAAAGGGGTG CAGCTGGGGG AGCTGTGGG AGRACAGGGT ACACAGCATC 1890
 GGGAAACCGAG CGCCGCGCTCG CTGACGTCCA GACTCCCGT CATCCCCAAG CCTGACGGGC TGCGCCCGAT 1960
 TGTGAACATG GACTACGTG TGGGAGCCAG AACGTTCCGC AGAGAAAAGA GCGGGCAGGC TCTCACCTCG 2030
 AGGGTGAAGG CACTGTTCAG CGTCTCAAC TAGGAGGGGG CGGGGGCCCC CGGCTCTCTG GCGCCTCTG 2100
 TGCTGGCCCT GGACGATATC CACAGGGCT CGGCGCACCTT CGTGTGCGGT GTGGGGGCCAGG AGGACCCGCC 2170
 GCGCTGACTG TACTTTGTCA AGGTGTATG GACGGGGCGG TACGACACCA TCCCCCAGGA CACGCTCACG 2240
 GAGGTGTCTCG CGACGATCAT CAAACCCCG AACACGCTAC GCGTGGCTCG GTATGGCGTG GTTCAGAAGG 2310
 CGCCCCATCG GCACGTCGCG AAGGCTCTCA AGAGCCACGT CTCTACCTTG ACAGACCTCC ACCCGTACAT 2380
 CGCACAGTTC GTGGCTCACCC TGCAGGAGAC CAGCCCGCTG AGGGATGCCG TCGTCACTGA CGAGAGCTCC 2450
 TCCCTGAATG AGGCCAGCAG TGGCTCTTC GACGTCTCAT TGTGGCACCCAG GCGTGGCGCA 2520
 TCAAGGGCAA GTCTGACATGC CAGTGGCCAGG GGATCCCCCA GCGCTCACAT CTCTCTACCG TGCTCTGCAG 2590
 CCTGTGCTAC CGCGCACATGG AGAACAAAGCT GTTGGGGGG ATTGGGGGG AGGGGGTGTG CCGCGCTTTG 2660
 GTGCGATGATT TCTTGTGGT GACACCTCTAC CTCAACCCAG CGAAAACCTT CCTCAGGACC CTGGTCCGAG 2730
 GTGTCCCTGA GTATGGCTGC GTGGTGAAC TCGGGAAAGAC AGTGGTGAAC TTCCCTGTAG AAGACGGAGGC 2800
 CCTGGGTGGC ACGGCTTTG TTCAGATGCC GCGCCACGGG CTATTCCTCTG GTGGGGCCCT GCTGCTGGAT 2870
 ACCCGGACCC TGGAGGTGCA GAGCGACTAC TCCAGCTATG CCGGGACCTC CATCAGACCC AGTCTCACCT 2940
 TCAACCCGCGG CTTCAGGGT GGGAGGAACAA TGCGTCGCAA ACTCTTTGGG GTCTGGGGC TGAGATGTCA 3010
 CAGCGCTGTTT CTGGATTGTC AGGTGAACAG CCTCCAGACG GTGTGCAACCA ACATCTACAA GTCCTCTCTG 3080
 CTGCAGGGCTG ACAGGTTTCA CGCATGTGTG CTGCAAGCTCC CATTCTCATCA GCAAGTTGGG AAGAACCCCCA 3150
 CATTTTCTCT GCGCGTCATC TCTGACACGG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACGCAGG 3220
 GATGTCGCTG GGGGGCGAGG GCGCTCTGGC CGAGGGGGGG AGGGGGCCCT GCGAGGAGG TGCACTGGCT 3290
 GCATTCCTGC TCAAGGTGAC TCGACACCGT GTCACTTACG TGCACTCTG GGGGTCACTC AGGACAGCCC 3360
 AGACCCAGCT GAGTCGGAG CGCCGGGGGA CGACGCTGAC TGCCCTGGAG GCGCAGCCA ACCCGGGCACT 3430
 GCCCTCAGAC TTCAAGACCA TCTGGACTG ATGGCCACCC GCGCCACAGCC AGGGGGAGG CAGACACCG 3500
 CAGCCCTGTC ACGGCGGGCT CTACGTCCAC GGGAGGGAGG GGCGGGCCAC ACCGGGGCC GCACCGCTGG 3570
 GAGTCGAGG CTCAGTGTAC TGTGGGGGG AGGGCTGCTG TGCCCTGCG AGGTGTAGTG TCCGGCTGAG 3640
 GCCTGAGCGA GTGTCCAGCC AAGGGCTGAG TGTCAGCAC ACCTGGGTC TTCACTTCCC CACAGGCTGG 3710
 CGCTCGGCTC CACCCCCAGGG CCAGCTTTG CTCACCAAGGA GCGGGCTTC CACTCCCCAC ATAGGAATAG 3780
 TCCATCCCCA GATTCGCCAT TGTTCACCCC TGCGCTGCG CTCCCTTGCC TTCCACCCCCC ACCATCCAGG 3850
 TGGAGACCCCT GAGAAGGAGC CTGGGAGCTC TGGGAATTTG GAGTGAACAA AGGTGTGCCCC TTGACACAGG 3920
 CGAGGACCCCT GCAACTGGAT GGGGGCTCTG GTGGGTCAA AAGGGGGAG GTGCTGTGGG AGTAAATAC 3990
 TGAATATATG AGTCTTCTAC TTTTGAAGAA AAAAAAAA AAAAAAAA AA 4042

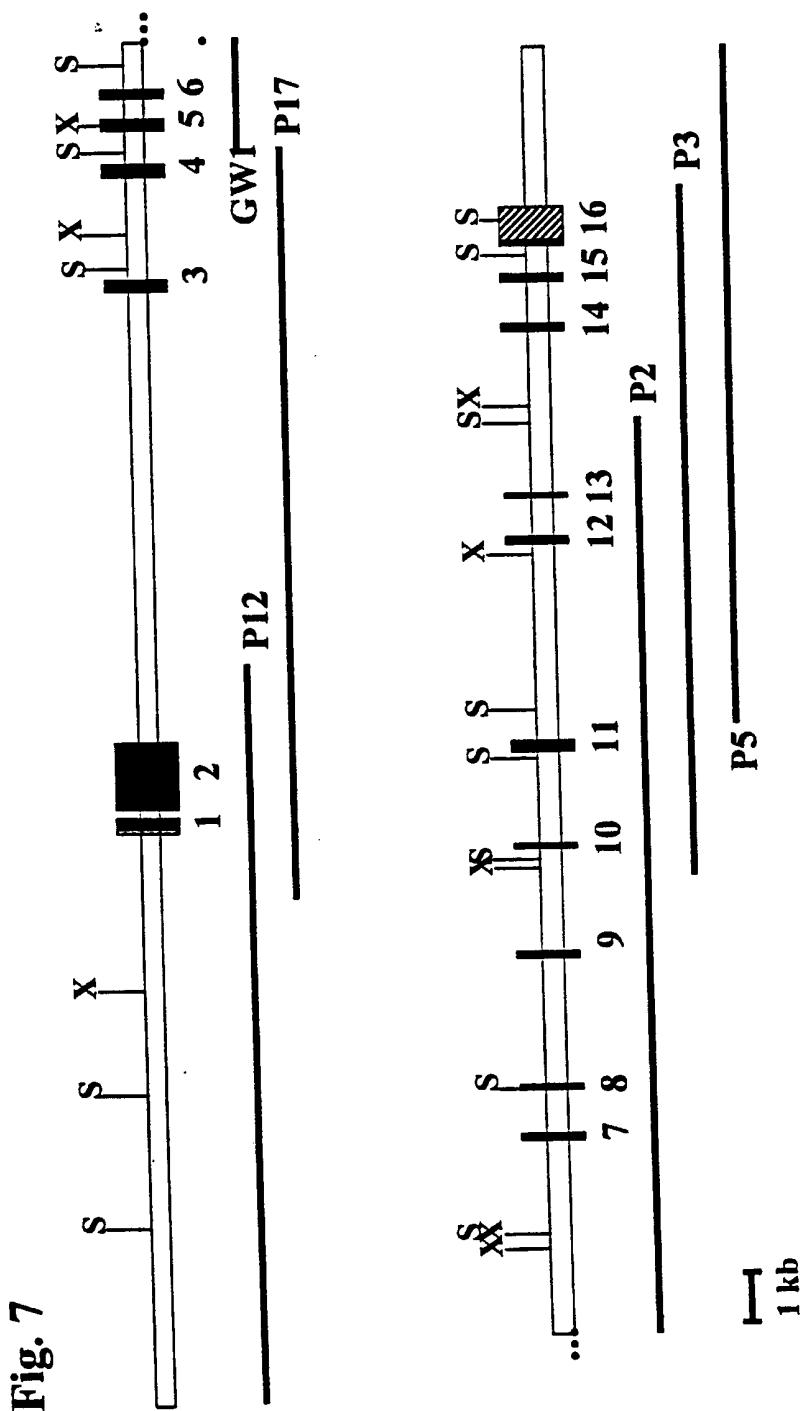
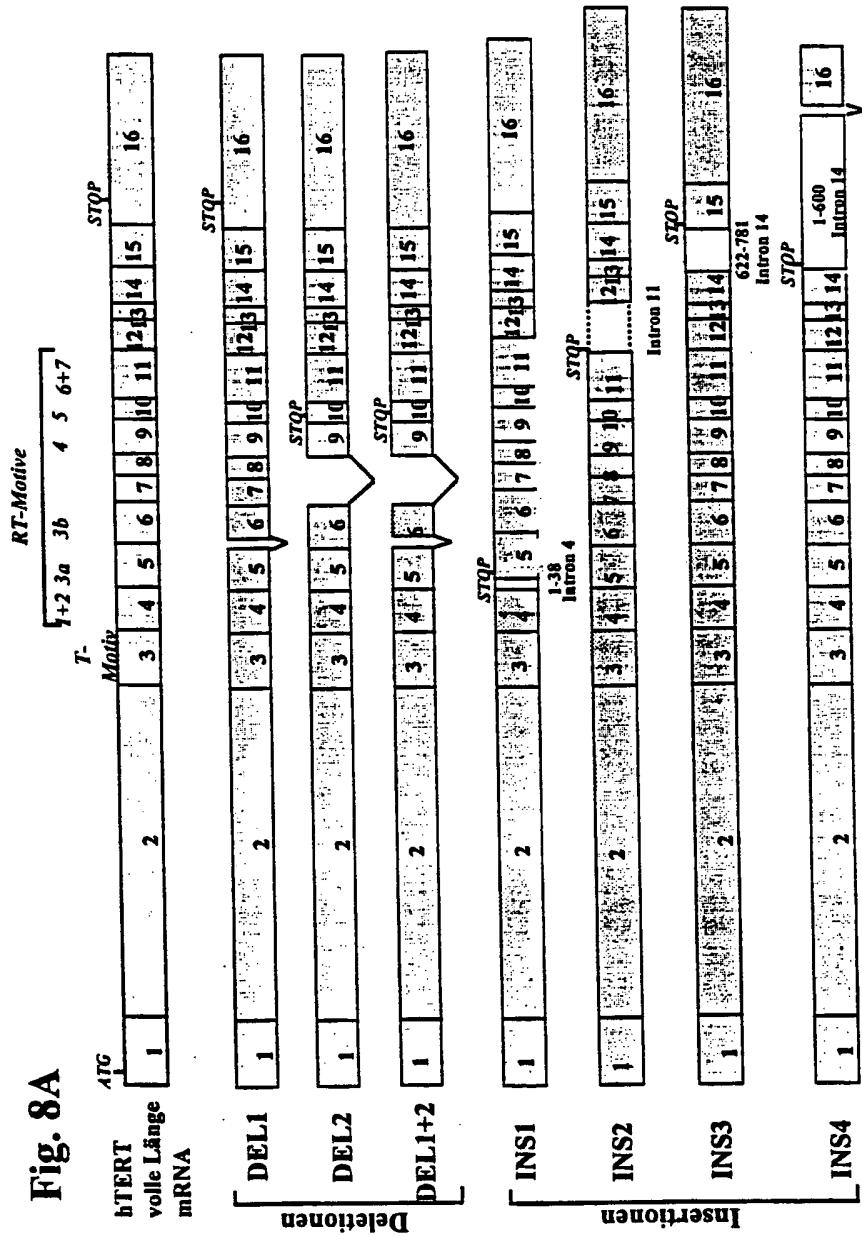


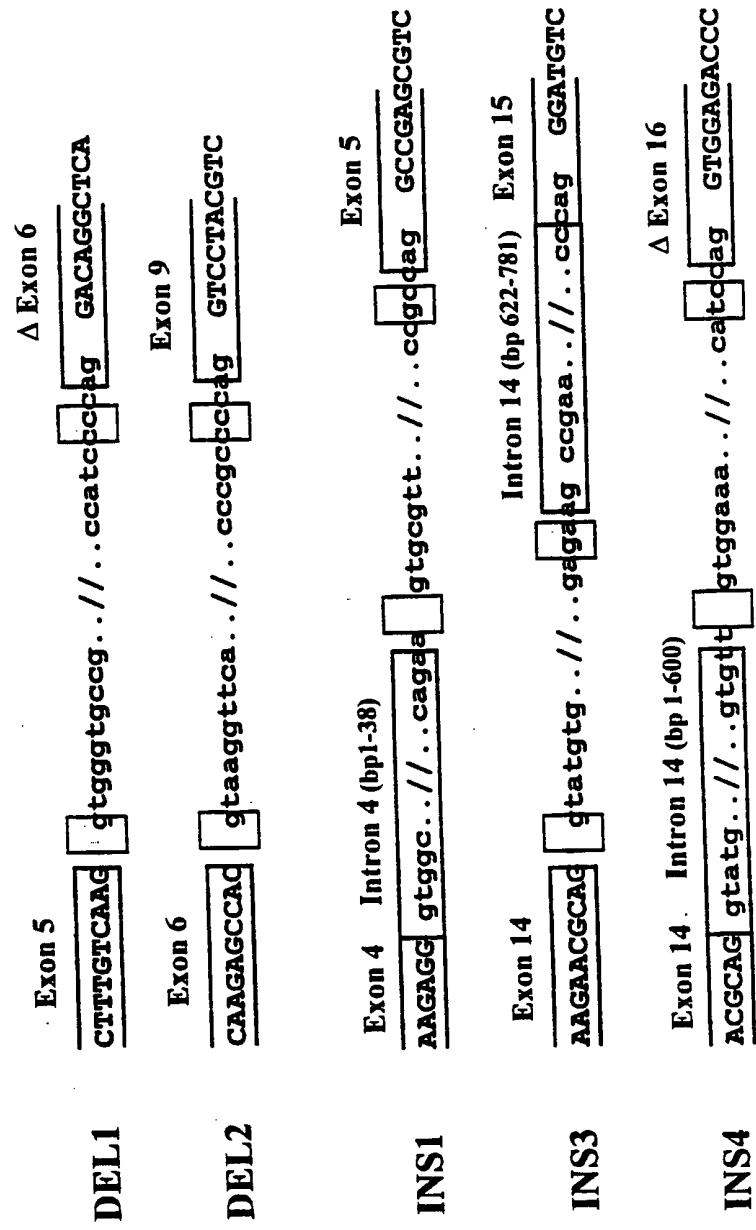
Fig. 7

Fig. 8A



ERSATZBLATT (REGEL 26)

Fig. 8B



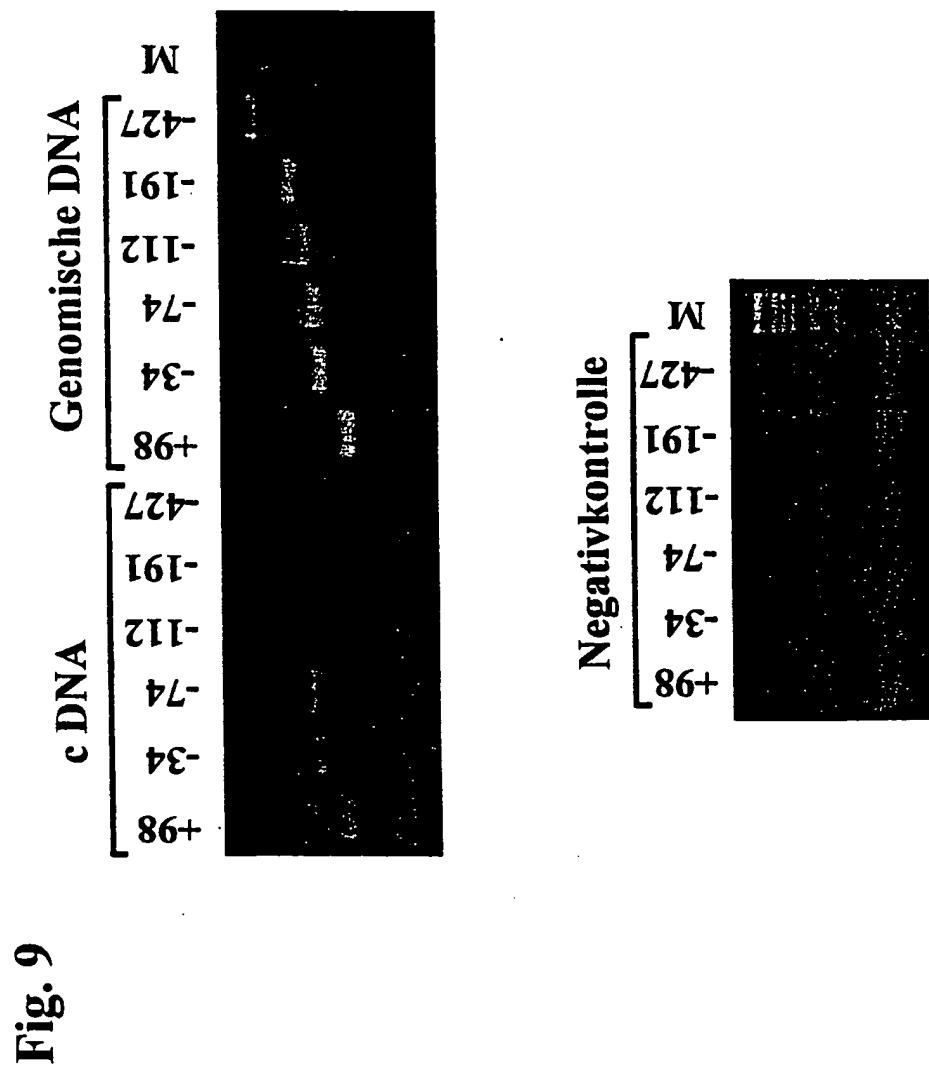


Fig. 9

Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
 ATGAGACCT GTCTCAAAAA AAAAAAAA AATTGAAAATA ATATAAGCA TCTTCTCTGG CCACAGTGGA -11134
 ACAAAACAG AATCAACAA CAAGAGGAAT TTGAAAACAT ATACAAACAC ATGAAAATTA AACAATATAC -11064
 TTCTGAATGA CCAGTGAGTC AATGAGAAA TTAAAAGGA AATTGAAAATA TTATTTAAG CAAATGATAA -10994
 CGGAAACATA ACCTCTCAA ACCCACGGTA TACAGCAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924
 AGCAGCTACA TCAAAAAGT AGAAAAGCCA GGGCGAGTGG CTCATGCCCTG TAATCCCAGC ACTTTGGGAG -10854
 GCCAAGGGCG GCAGATCGCC TGAGGTCAGG AGTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTG -10784
 CTACTAAAAA TACAAAATTA GCTGGCATG GTGGCAGATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG -10714
 GCAGGATAAC CGCTGAAACC CAGGAGGTGG AGGTTGCGGT GAGCGGGAT TGCGCCATTG GACTCCAGCC -10644
 TGGGTAACAA GAGTGAACCC CTGTCCTCAA AAAAAAAA AAGTAGAAA ACTTAAAAT ACAACCTAAT -10574
 GATGCACTT AAGAACTAG AAACTCAAG CCAAAACTAA CCTTAAATG GTAAAAGAAA AGAAATAATA -10504
 AAGATCAGAG CAGAAAATAA TGAAAATGAA AGATAACAAT ACAAAAGATC AACAAATTA AAAGTTGGTT -10434
 TTTTGAAGAAG ATAAACAAA TTGACAAACC TTGCCCCAGA CTAAGAAAAGGAGAAG ACCTAAATAA -10364
 ATAAAGTCAG AGATGAAAAGGAGACATTA CAACTGATAC CACAGAAAATTA CAAAGGATCA CTAGAGGCTA -10294
 CTATGAGCAA CTGTCACATA ATAAATTGAA AACCTAGAA AATTCCTAGA TGCACTAAC -10224
 CTACCAAGAT TGAACCATGA AGAAAATCAA ACCCCAAACA GACCAATAAC AATAATGGGA TTAAAGCCAT -10154
 AATAAAAAGT CTCCTAGCAA AGAGAAGGCC AGGACCAAT GGCTCCCTG CTGGATTAA CCAATCATT -10084
 AAAGAAGAAT GAATTCCAAT CCTACTCAAAT CTATTCTGAA AAATAGAGGA AGAAATACTT CCAAACCTAT -10014
 TCTACATGGC CAGTATTACG CTGATTCCAA ACCGACAGAA AACACATCA AACAACAAACA AACAAAAAAA -9944
 CAGAAAGAGA GAAACTACA GGCAATATC CCGTGAATGAA ACTGATACAA AATCCTCAA CAAACACTA -9874
 GCAACACAA TAAACAAACA CCTCTGAAAG ATCATTCAATT GTGATCAAGT GGGATTATT CCAGGGATGG -9804
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAACACTA -9734
 TATGATTATT TCACATTATG CAGAAAAAGC ATTTGATAAA ATTCATGCACC CTTCATGATA AAAACCCCTCA -9664
 AAAAACCCAGG TATACAGAA ACATACAGGC CAGGGACAGTGG GGCTCACACC TGCACTCCCA GCACCTCTGG -9594
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTGAG ACTAGCTGG GCAACAAAT GAGACCTGGT -9524
 CTACAAAAAA CTTTTTAAA AATTTAGCCA GGCATGATGG CATATGCCCTG TAGTCCCAGC TAGTCTGGAG -9454
 GCTGAGGTTG GAGAATCACT TAACTCTAGG AGGTCGAGGC TGCACTGAGC CATGAACATG TCACTGTA -9384
 CCAGCCTAGA CAACAGAACAA AGACCCCACT GAATAAGAGAAG AGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAGTGG AGGGAGGAGG GGAAGGGAGA GAGGAGAAG -9244
 AAGAAACATA TTCAACATA ATAAAGGCC TATATGAGAC ACCGAGGTAG TATTATGAGG AAAACTGAA -9174
 AGCCTTCTCT CTAAGATCTG GAAAATGACA AGGGCCCACT TTCACTACTG TGATTCAACA TAGTACTAGA -9104
 AGTCCTAGT AGAGCAATCA GATAAGAGAA AGAAAATAAA GGCATCCAA CTGGAAGAGA AGAAGCTCAA -9034
 TTATCCCTGTG TGAGATGAT ATGATCTTAT ATCTGAAAAGA GACTTAAAGAC ACCACTAA AACTATTAGA -8964
 GCTGAAATTG GGTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTCTCTATA TTCCAACAGC -8894
 AAACAATCTG AAAAGAAAC CAAACAGCA GCTACAAATAA AAATTAACAA GCTAGGAATT AACCAAAGAA -8824
 GTGAAAGATC TCTACATGA AAACATATAAA ATGTTGATAA AAGAAATGAA AGAGGGCACA AAAAAGAAA -8754
 AGATATTCCA TGTTCATAGA TTGGAAGAAT AAATACTGTT AAAATGTCCTA TACTACCCAA AGCAATTTCAC -8684
 AAATCAATG CAATCCCTAT TAAACATACTA ATGACCTTCT TCAACAGAAAGA AGAAGAACAA ATTCTAAGAT -8614
 TTGACAGAGA CCACAAAGA CCCAGAATAG CCAAAGCTAT CTCGACAAA AAGAACACAA CTGGAAGCAT -8544
 CACATTACCT GACTCTCAAAT TATACATACAA AGCTATAGTA ACCCAAACCA CATGGTACTG GCATAAAAC -8474
 AGATGAGACA TGACCCAGAG GAACAGAATAA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACCTCA -8404
 TTTTGACAAAG AGGTGCCAAG AACACACTTT GGGGAAAGA TAATCTCTC ATAATAGGT GCTGGAGGA -8334
 CTGGATATCC ATGTCACAA TAAACATACTA AGAACCTGT CTCTCACCAT ATACAAAGC AAATCAAAT -8264
 GGATGAAAGG CTAAATCTA AAACCTCAA CTTTGCACAT ACTAAAGAGA AACACCCGGAG AAACCTCTCA -8194
 GGACATTGGA GTGGCCAAAG ACTTCTTGAG TAATTCCTG CAGGCACAGG CAACCAAAGC AAAACACAGC -8124
 AAATGGGATC ATATCAAGTT AAAAGCTTC TGCCCAAGCAA AGGAAACAT CAACAAAGAG AAGAGACAAAC -8054
 CCACAGAAATG GGAGAATATAA TTGCAACACTT ATTCATCTAA CAACGAAATAA ATAACCAAGTA TATATAAGGA -7984
 GCTCAAACACTA CTCTATAAGA AAACACCTA ATAAGCTGAT TTCAAAAT AAGCAGAAAGA TCTGGTAGA -7914
 CATTCTCAA AATAAGTCACAAATGCA AACAGCATG GAACAGCATC TGAAAATGTC CTCACACCA CTGATCATCA -7844
 GAGAAATGCA AATCAAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAA ATGGCTTTA TTCAAAGAC -7774
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGAC ACTGTTGTTG GGAATGGAAA -7704
 TTGCTACAC TATGGAGAAC AGTTGAAAG TTCCCTCAA AACTAAATTA AAAGCTACCA TACAGCAATC -7634
 CCATTGCTAG GTATATACCA CAAACAGGAA ATCAGTGTAA TCAACAAAGCT ATCTCCACTC CCACATTTCAC -7564
 TGCAGCACTG TTTCATAGCAG CCAAGGTTTG GAACACCT CAGTCCTAT CAACAGACGA ATGGAAAAAG -7494
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAAGCA TAAAAAGAA TGAGATCTG TGAGATCTGAA -7424
 CAGCATGGGG GGCACATGTC ACTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354
 CTCCCTTACT TGTGGGAGCA AAAATTTAAA CAATTGACAT AGAAATAGAG GAGAATGGT GTTCTAGAGG -7284
 GGTGGGGGAC AGGGTCAAG GACTCAACAA TAATTATTG TATGTTTAA AATAACTAA AGAGTATAAT -7214
 TGGGTTGTT GTACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTTAC CCTGATGTGA -7144
 TTATTACACA TTGTATGCCT GTATCAAAT ATCTCATGTA TGCTATAGAT ATAAACCCCTA CTATATTAA -7074
 AATTTAAATTT TTAATGGCCA GGCACTGGTG CTCATGTCG TAATCCCAGC ACTTTGGGAG GCGGAGGCGG -7004
 GTGGATCACC TGAGGTCAGG AGTTGAAAC CAGCTGGCC ACCATGATGA AACCTGTCT CTACTAAAGA -6934
 TACCAAAAT AGCAGGGCGT GGTGGCACAT ACCTGTAGTC CCAAACACTC AGGAGGCTGA GACAGGAGAA -6864
 TTGCTGAAAC CTGGGAGGGCG GAGGTTGCAAG TGAGCCAGA TCATGCCACT GCACAGCAGC CTGGGTGACA -6794
 GAGCAAGACT CCATCTCAA ACAAACAA AAAAAGAG ATTAAAATG TAATTTTTAT GTACCGTATA -6724
 AATATATAC TCACTATATT AGAAGTTAAA AATTTAAACCA ATTATAAAG GAAATTAACC ACTTAATCTA -6654
 AATAAAGAAC AATGATGTTG GGGTTCTAG CTTCTGAAGA AGTAAAAGTT ATGCCACGA TGGCAGAAAT -6584

Fig. 10

GTGAGGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTAACCC -6514
 AAAGACAGGC TGGGAGAAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAGG -6444
 TAATCTCTAT TAATACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATGGCA CGTCCTGATCA -6374
 CACCGCTCTC TCATTCAAGG TGCTTTTTT CTTGTGTGCT TGGAGATTT CGATTGTGTT TTCGTGTTG -6304
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAATGGT GGTGATTTC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGGGGCTC TGTTGACCTG AGCCRCTTC ATCTTCAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGGCTGTG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
 CGTTGGTGAAG CAGGGCATGA ACTGCCCTTA TTTACGCTT GCAAAGATG CTCCTGGATAC CATCTGGAAA -6024
 AGGGGGCCAG CGGGATGCA AGGAGTCAGA AGCCTCTGC TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954
 CCCGGGCTG TGCCAGAGGG AGAGGAGTCAGA AGGCACCTCG AAGTATGGC TAAATCTTT TTTCACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAAG CTTGAGTTAG GTGCTCTT TAAACAGAA AGTCATGGAA -5814
 GCACCCCTCT CAAGGGAAAA CCAGACGCC GCTCTGGGT CATTACCTC TTTCCTCTT CCCTCTCTT -5744
 CCCCTCGGGT TTCTGATCGG GACAGAGTGA CCCCTCTGGA GCTTCTCGGA GCCCCGCTG AGGACCCCT -5674
 TGCAAAGGGC TCCACAGACC CCCGCCCTGG AGAGGAGGT CTGAGCCTGG CTTAAATACA AACTGGGATG -5604
 TGCGCTGGGG CGGACAGCGA CGGGGGGATT CAAAGACTTA ATTCCTAGG TAAATTCAC TTTCCACAT -5534
 CCGAATGGAT TTGGATTTA TCTTAATATT TTCTTAAATT TCATCAATAA ACATTCAAGG CTGCAAGAAAT -5464
 CCAAAGGCCTG AAAACAGGA CTGAGCTATG TTTGCAAGG TCCAAGGACT TAATACCAT GTTCAGAGGG -5394
 ATTTTCGCCC CTAGTACTT TTATTGTTG TTCATAAGGT GCTTGTAGGT GCAAGGGAAA GTACACGAGG -5324
 AGAGGCTGG CGGGCAGGG TATGAGCAG GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
 ACACAGGAC CACTGACCGT CCTCCCTGG AGCTGCACCA TTGGGCAAGG CGAAGGGCCG CACGCTGCGT -5184
 GTGACTCAGG ACCCCATACC GGCTTCTGG GCCCACCCAG ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114
 CCGTGGAAAC GAACATGACC TTGCTCTGCC TGCTTCCCTG GGTGGTCAAG GGGTAATGAA GTGGTGTGCA -5044
 GGAATGGCC ATGTAATTA CACAGCTCTG CTGATGGGA CGCTTCCCTC CATCATATT CATCTTCACC -4974
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGAAAGCCATG CAAACATCAG TACAAACACC -4904
 ACTCTTTAC TAGGCCACA GAGCACGGSC CACACCCCTG ATATTAATAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTCAGCC ACCAGGCTGG GGTGACAACA GGGGCTGAAC AGTCTGTCC TCTAGACTAG TAGACCCCTGG -4764
 CAGGGACTCC CCCAGACTTCT AGGGCTCTGG TGTCTCTCC CGAGGGGCCG ATCTGCCCTG GAGACTCAGC -4694
 CTGGGGTGC ACATGAGGG CAGGGCTCTG TCCACACCC CGCCCTCCAG GCCTCAGCTT CTCCAGCAGC -4624
 TTCTTAAACC CTGGGGTGGGC CGTGTCTGG ACCTGACTCG TCCCTCTGTC CACTGTGCT TGTCTCAGCG -4554
 ACGTAGCTCG CACGGTTCTC CCTCATGGGG GGTGTCTGTC TCCCTCCCA ACACTCACAT GCGTTGAAGG -4484
 GAGGAGATTG TGCCCTCTCC AGACTGGCTC CTCTGAGCCT GAACTGGCT CCGGGGGCCC GATGCAGGTT -4414
 CCTGGGTCTC GGCTGCACCC TGACCTCTCAT TTCCAGGGCC TCCCTCTGTC CTGTCATCTG CCGGGGCTG -4344
 CCGGTGTGTT CTCTGTTCTC TGTCCTT GGGGTTTTA TAGGCATAGG ACGGGGGGCT GTGGGCTCTG GGAAATGCAA CATTGGGTG -4204
 TGAAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGGCA CAGGGCTGGG GATGGAGCCC CCGGCAAGGG -4134
 CCCGCCCTTC TCTGCCCAAGC ACTTTCTGC CCCCTCTCTT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCAA AAGACCCAGC ATTGGCACCC CTGGACATTG GCCCCACAGC CCTGGAAATT -3994

~Mu~

CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCGG CTGTTTATT TTAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAAATGTCC TTTAACARAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTCC TCACAGTGAAG GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTAAAAGG AAAGAATTTTACCCATGGC -3714
 AGGGGACTGG TTAGGGGGT TAAGGAGGT GGGGGCGGA GCTGGGGGCT ACTGCACGCA CCTTTACTA -3644
 AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCACTGTTG GGAGACTAAC CATAGGGAG TGGGATGGG -3574
 GGAACCCGGG GGCTGTGCCA TCTTGCCT GCCCAGGTG CCGGGCAGG ATAATGCTCT AGAGATGCC -3504
 ACGTCTCTGAT TCCCCAACAC CTGTGGACAG AACCCGGCCG GCCCAGGGC CTTTGCAAGG GTGATCTCCG -3434
 TGAGGACCTC GAGGTCTGGG ATCTCTGGG ACTACCTGCA GGCCCCAAAAA GAAATCCAGG GGTCTGGGA -3364
 AGAGGGGGC AGGGGGTCA GAGGGGGCA GCGCTCAGG GATGGAGGCA GTCAGTCTG GGCTGAAAAG -3294
 GGAGGGAGGG CCTCGAGGCC AGGGCTGCCA GCGCTCAGG AAGCTGGAAA AAGCGGGGAA GGGACCCCTCC -3224
 ACGGAGCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCG ATCGTGGACC TCCGGCTCC -3154
 GTGCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTTGGGG ATTTCAGGAA GCAACAGGAA -3084
 ACCCATGAC TGTGAATCTA GGATTATTC AAAACAAAGG TTTCAGAAA CATCCAAGGA CAGGGCTGAA -3014
 CTGCTCTGGG GCAAGGGCAG GCGCAGGAGC AGTGTATTG TTAGTATTG TTATTTATT TACTTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTTGCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCACTC -2874
 CGTCCTCTGG GTTCAAGCAA TTCTCGTGC TCAGCCTCCC AAGTAGCTGG GATTCAGGC GTGCACCA -2804
 ACACCCGGCT AATTTGTAT TTTTAGTGA GATGGGGTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC -2734
 CTGACCTCAG GTGATGCCG CACCTCAGGC TCCCAAAGT GCGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCCTATTTA ACCATTAA AACTTCCCTG GGCTCAAGTC ACACCCRCTG TAAGGAGGTT CATGGAGTTC -2594
 AATTCCCTT TTACTCAGGA GTTACCCCTCC TTGATATT TCTGTAATT TTCGTAGACT GGGGATACAC -2524
 CGTCCTCTGG CATATTCAAGA TTGTTCTGTGAA CCACCTGTA TCCCATGGG CCCACTGCAG GGGCAGCTGG -2454
 GAGGCTGCAG GCTTCAGGTC CAGGGGGGGT TGCCATCTGC CAGTAGAACAC CTGAGTGTAGA ATCAGGGCGC -2384
 AAGTGTGGAC ACTGTCTCTGA ATCTCAATGT CTCACTGTT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314
 CTCCCTACTCT ACTGGGATTG AGCCCCCTCC CTATCCCCC CCAGGGGAGG AGGAGTTCTC CTCACTCTG -2244
 TGGAGGAAGG AATGATACTT TGTTATTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGGTTG -2174
 TTTGTTTGTT GTTGGAGGGC GTTTCACTC TTGTTGCTCA GGCTGGAGGG AGTGCATGG CGCGATCTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGAATCTC CTGCTTCCGC CTCCCATTTG GCTGGGATTA -2034
 CAGGCACCCG CCACCATGCC CAGCTAATTG TTGTTATTT TAGTAGAGAC GGGGGTGGGT GGGGTTCAAC -1964

Fig. 10

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
 GATTACAGGT GTGAGGCCACC ATGCCAGCT CAGAATTAC TCTGTTAGA AACATCTGGG TCTGAGGTAG -1824
^{CAAT-Box}
 GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAGGCAA TGTAGAATT TTTTATTGT TGTTAGAAC -1754
 CTCTTGATGT TTTCACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCAACCCATA -1684
 ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTCTCG CCATGCACAT -1614
 GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCTCT TCTCTTCCCT CTTTTAAAAT TGTGTTTCT -1544
 ATGTTGGCTT CTCTGCAGAG AACCAGTGT AAGCTACAAT TAACTTTGT TGGAAACAAT TTTCCAACC -1474
^{Spl}
CCCCTTGCA CCTAGTGGCA GAGACAATTG ACAAAACACAG CCCTTTAAAA AGGCTTAGGG ATCAGTAAAGG -1404
 GGATTCTAG AAGAGCGACC TGTAAATCCTA AGTATTACAGA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334
 CCCCAGGGAG GGTGGAGGC CTGTTCAAT GCTAGCTCCA TAAATAAGC AATTCCTCC GGCAGTTCT -1264
 GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTTCACTGT TTGCGACCT CAGCTACAGC -1194
 ATCCCTGCAA GGCTCGGGA GACCCAGAAG TTTCTCGCC CTTAGATCC AAACCTGAGC AACCCGGAGT -1124
 CTGGATTCTT GGGAAAGTCCT CAGCTGTCTT GCGGGTTGTGC CGGGGCCCCA GGTCTGGAGG GGACCAAGTGG -1054
 CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984
 GCCTGGACCC CGAGGCTGCC CTCCACCCCTG TGCGGGCGGG ATGTGACCAAG ATGTTGGCCT CATCTGCCAG -914
 ACAGAGTGC CGGGGCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCCG CCGGTGCGCG GCCAGCAGGA -844
^{CCAC-Box}
^{Spl}
 GCCTGGCT CCATTCCCA CCCTTCTCG ACAGGGACCCG CCCGGTGGGT GATTAACAGA TTTGGGGTGG -774
 TTTGCTCATG GTGGGGACCC CTCGCCGCCT GAGAACCTGC AAAGAGAAAT GACGGGCCTG TGTCAAGGAG -704
 CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCGT GCCCCTCCAG GGAGCAATGC -634
^{AP-2}
 GTCCCTGGGT TCGCCCCCAG CGCGGTCTAC GCGCCTCCGT CCTCCCCCTTC ACGTCCGGCA TTGCGTGGTGC -564
 CCGGAGCCCG ACGGCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGGCAAAG -494
 GGTCCGGCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCCTCCCT CGGGTACCC CACAGCCTAG -424
^{Spl}
 GCCTGGCTGA CCTCTCTCCG CTGGGGCCCT CGCTGGCTC CCTGCACCCCT GGGAGCGCGA GCGGCCGCG -354
^{Spl}
GGCGGGAAAG CGGGGCCCAG ACCCCCCGGGT CCGCGGAG CAGCTGGCCT GTCGGGGCCA GGCGGGCTC -284
^{c-Myc}
 CCAGTGGATT CGGGGGCACA GACGGCCAGG ACCGGCTCC CCACGTGCG GAGGGACTGG GGACCCGGGC -214
^{Spl}
 ACCCGTCCCTG CCCCTTCACC TTCCAGCTCC GCCTCCCTCG CGCGGACCC GCGCGTCCC GACCCCTCCC -144
^{Spl}
 GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCCTT TCCGCGGCCCC CGCCCCCTCC -74
^{c-Myc}
 TCGCGGCCG AGTTTCAGGC AGCGCTGCCT CCTGCTGCCC ACGTGGGAAG CCCTGGCCCC GGCCACCCCCC -4
 GCGATG

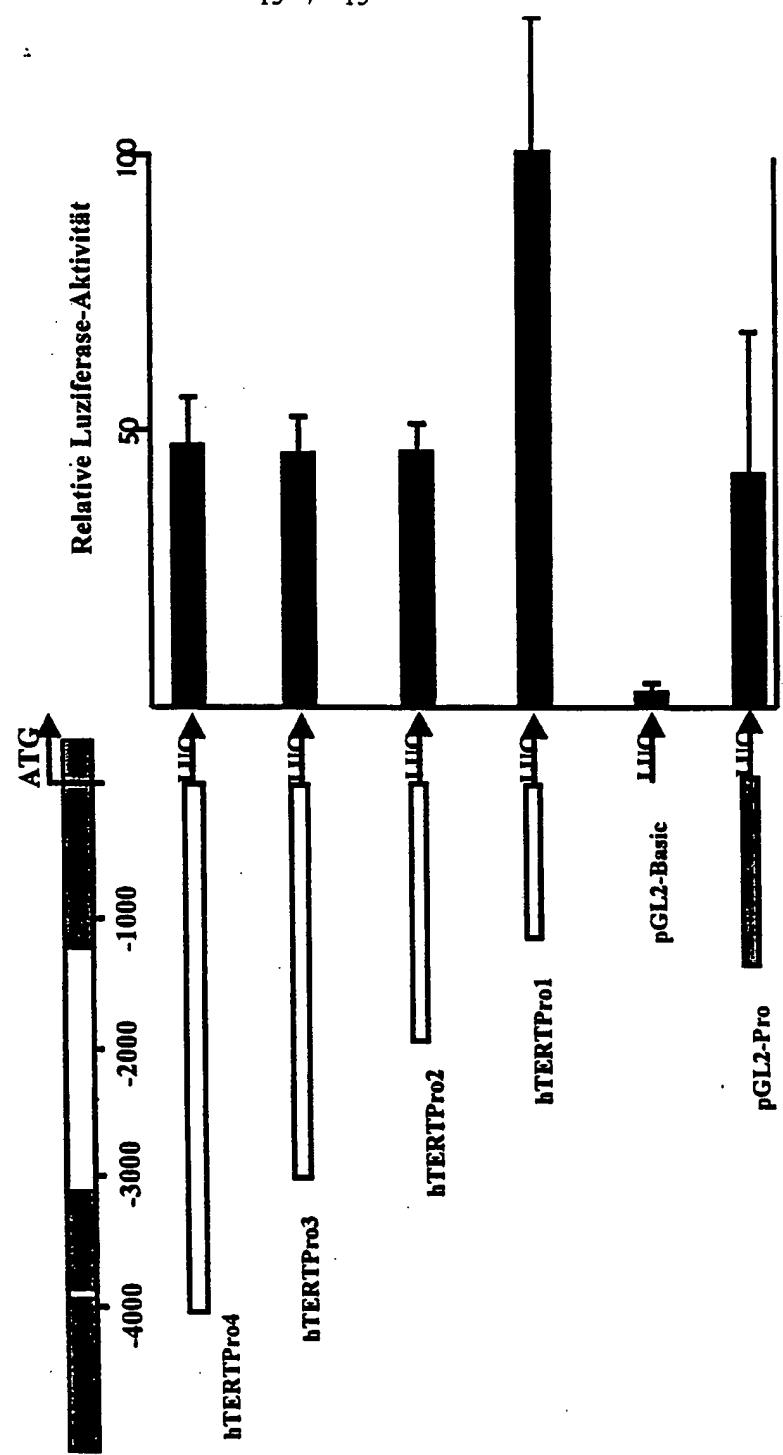


Fig.: 11

SEQUENZPROTOKOLL

<110> Bayer AG

5	gtggcgctgc	ccctgagccg	gagccggacgc	ccgttgggca	ggggctctgg	gcccccccg	840
	gcaggacgcg	tggaccgagt	gaccgtgtt	tctgtgtgtt	gtcacctgc	agaccccg	900
	aagaagccac	cttttggag	ggtgcgtct	ctggcacgcg	ccactccac	ccatccgtgg	960
	gcccggacca	ccacgcggc	ccccccatcca	catcgccggc	accacgtccc	tgggacacgc	1020
	cttgcgtccc	ggtgtacggc	gagaccaage	acttcttc	cttccttc	gacaaggagc	1080
	agtcgtggcc	ctcttctta	ctcagcttc	tgagggccag	cctgtactgg	gtctcgaggc	1140
10	tctgtggagac	catctttctg	ggttccaggc	cctggatgcc	agggactccc	cgcaagggtc	1200
	cccgcctgcc	ccagcgctac	tggcaaatgc	ggcccctgtt	tctggagctg	cttggaaacc	1260
	acgcgcgtg	ccccatacggg	gtgtcttc	agacgcactg	cccgtgcgca	gtctcggtca	1320
	cccccgacgc	cggtgtctgg	gccccggaga	agccccccgg	ctctgtgtgc	gccccccagg	1380
	aggaggacac	agacccccgt	cgccctgtgt	agtcgtctcc	ccagcacage	agcccccttc	1440
15	agggtgtacgg	cttcgtgtgg	gtctgtgtcc	gccccgtgt	gccccccaggc	cttcgtgggt	1500
	ccaggcaca	cgaacgcgc	ttcttc	acaccaaga	gttcatctcc	ctgggaaagc	1560
	atgccaagct	ctcgctgcag	gagctgacgt	ggaagatgag	cgtrggggac	tgctgtttgc	1620
	tgccgcaggag	ccccagggtt	ggctgtgttc	cgcccgccaga	gcacccgtcg	cgtrggaggac	1680
	tcctggccaa	gttcctgcac	tggctgtatg	gtgtgtatgt	cgtrggatgt	ctcagggttt	1740
20	tctttttatgt	cacggagacc	acgtttcaaa	agaacaggct	ttttttctac	cgaaagagatg	1800
	tctggagcaa	gttgc当地	attggaaatca	gacagcacct	gaagagggtg	cagctgcggg	1860
	agctgtcgga	agcagaggac	aggcagcatc	ggaaagccag	gccccccctg	ctgacgtcca	1920
	gactccgtt	catccccaa	cctgacgggc	tgcggccgat	tgtgaatcatg	gactactcg	1980
	tgggaggccag	aaacgttccgc	agagaaaaga	ggggccggcg	tctcaccc	agggttaagg	2040
25	cactgttca	cgtgtcaac	taacggatcc	cgccggccgc	cgccctccgt	ggcgcctctg	2100
	tgctgggtct	ggacgatata	caacgggctt	ggcgcaccc	cgtgtgcgt	tgccggggcc	2160
	aggacccggc	gcctgagctg	tactttgtca	aggtggatgt	gacggggcgc	taegacacca	2220
	tccccccagg	caggctca	gaggtcatcg	ccagcatcat	caaaccccg	aaacacgtact	2280
	gcgtgcgtcg	gtatgcgt	gtccagaagg	cgccccatgg	gcacgcgtcg	aaaggccctca	2340
30	agacccacgt	ctctaccc	tttgcgtatc	acagacccct	gacgcacatc	ctcaccac	2400
	tgcaggagac	cacggccgt	aggatgtcc	tgcgtatcg	cgacgtctcc	tccctgtatg	2460
	aggccacgcag	tggtcttc	gacgtcttc	ttacgttcat	gtccacccac	gcccgtgc	2520
	tcaggggcaa	gtcctacgtc	cagtgtccagg	ggatcccgca	gggtctccatc	ctctccacgc	2580
	tgcctctgcag	cctgtgtca	ggcgacatcg	agaacaaatct	gttttgcgggg	atttcggccgg	2640
35	acgggtgtct	cctgtgttt	gtggatgtt	tcttgcgtgt	gacacccatc	ctcaccac	2700
	cggaaaaccc	cctcaggacc	ctgtgtcccg	gtgtgtccgt	agggtgtcc	tggtgttaact	2760
	tgcggaaagac	agtgtgtac	ttccctgtatg	aagacggac	cctgggttgc	acgggttttgc	2820
40	ttcagatgcc	ggcccacggc	ctatcccc	ggtgcggct	gtgtgttgc	acccggaccc	2880
	tggagggtgc	gagcgactac	tccagatct	ccccggatcc	catcgacggc	agtctccatc	2940
	tcaacccgc	cttcacgggt	ggggagaaca	tgcgtgc	actttttgg	gttttgcggc	3000
	tgaagtgtca	cagccgtt	ctggat	tttgcgtatc	cccccacgg	tggtgtccacca	3060
	acatccatca	gatcccc	ctgtgtccgt	acaggtttca	cgatgtgt	tgctgtcc	3120
45	catttcatca	gcaagtttgg	aagaacccca	catttttct	gcccgttc	tctgacaccc	3180
	cctccctctg	ctactccatc	ctgaaagccca	agaacgcagg	gtgtgtcc	ggggccaa	3240
	gcccggccgg	cccttc	tccggaggcc	tgcgtgt	gtccacccaa	gtatccctgc	3300
	tcaagctgac	tcgacacccgt	gtcacccatc	ttccgtactc	gggggtc	aggacacgc	3360
	agacgcgtgt	gatgtgtcc	cacccgggg	cgatgtgtcc	ccctgggtgg	gcccgcacca	3420
50	acccggca	gccttc	ccatggac	ttcaagacca	tcctggactg	atggccaccc	3480
	aggccgagag	cacaccc	cacccgt	acgcggggct	ctacgttcc	gggaggagg	3540
	gcccggccac	acccggcc	gcacccgt	gatgtgtcc	tcgtgtgt	tggttgcggc	3600
	aggcctgtat	gtccggctg	aggctgtat	tccgggtgt	gctgtgtcc	gtgtccaccc	3660
	aagggtgtat	tgccgtcc	acccgtcc	ttcaccc	ccacagggtt	cgctcggtcc	3720
55	cacccca	ccagttt	ctcaccagg	gcccggctt	caetccccc	ataggaatag	3780
	ccatccca	gattcgccat	tgttcc	tcgcctgtcc	ctcccttgc	ttccacccccc	3840
	accatccagg	tggagaccc	gagaaggac	ctgggagctc	tgggat	gatgtgtccaa	3900
	aggtgtgtcc	tgttcc	acccgt	cgacccgt	gcccccc	gggggttccct	3960
	trggggggag	gtgtgtgt	agtaaaaat	tgaatata	atgttttgc	ttttgaaaaaa	4020
	aaaaaaaaaa	aaaaaaaaaa	aaaaaaa	aa			4042
60	<210> 3						
	<211> 11276						
	<212> DNA						
	<213> Homo sapiens						
65	<400> 3						
	acttgcgtcc	aaagatgtca	ggctacggtg	agccatgatt	gcaacaccac	acgcggaccc	60
	tgggtgcaga	atgagaccc	gtctcaaaaa	aaaaaaaaaa	aattgtaaata	atataaagca	120
	cttcttc	ccacacgttca	acaaaaacca	aaatcaacaa	caagggaaat	tttggaaaaact	180

atacaaaacac atgaaaattt aacaatatac ttctgaatga ccagtggatc aatgaagaaa 240
 taaaaaaggaa aattgaaaaa ttatattaag caaatgataa cggaaacata acctctcaa 300
 acccacggta tacagcaaaa gcagtctaa gaaggaaattt tatactata agcagctaca 360
 5 tcaaaaaagt agaaaagcca ggcgcagttt ctatggctg taatcccacg accttggag 420
 gccaaggccg gcatatcgcc tgaggctagg agttcgacg cagctgacc aacacagaga 480
 aaccttgcg ctactaaaaa tacaaaattt gctggccatg gtggcacatg cctgtatcc 540
 cagctactcg ggaggcttag gcaggataac cgcttgaacc caggaggtag aggttgcgg 600
 gagccggat tgcgcattt gactccagcc tggtaacaa gatgtaaaacc ctgtctcaag 660
 aaaaaaaaaa aagttagaaaa actttaaaaat acaacatatac gatgcacccaa aagaacttag 720
 10 aaaaaggcaaa gcaacttattt cttttttttt gtaaaaaggaa aaaaataataa aagatcgag 780
 cagaaatataa taaaactgaa agataacaat caaaaggatc aacaaaattt aagatgggtt 840
 ttttggaaatg ataaacaaaacc ttggcccaaa ctaagaaaaa aggaaagaag 900
 accttaataa ataaagttag agatgaaaaa agagacattttaa aactgtatac cacagaaattt 960
 caaaggatca ctatggatca ctatggatca ataaatttggaa aacacttagaa 1020
 15 aaaaatatataa aatccctttaga tgcatacaac ctatggatca tgaaccatgaa aaaaatccaa 1080
 agcccaaaaac gccaataac aataatggaa tttttttttt aataaaaggatc ctccttagca 1140
 agagaagccc aggacccatg ggtttccctt ctggatattt ccaatctttt aaaaagaat 1200
 gaattccat cttttttttt ctatggatca aatagtaggaa aaaaatctttt ccaatcttcat 1260
 20 tctatcatggc cttttttttt ctatggatca aaccagacaa aacacatca aaaaacaaaca 1320
 aaaaaaaaaggaa cttttttttt cttttttttt ggtttccctt ctggatattt aatggatca 1380
 aatcccttcaaa cttttttttt cttttttttt tttttttttt aatggatca aatggatca 1440
 gtgtatcaatg gggatattt ccaggatggaa aatggatggatc aatggatca aatggatca 1500
 atgtatcaatg tttttttttt aatggatca aatggatca aatggatca aatggatca 1560
 25 cttttttttt cttttttttt aatggatca aatggatca aatggatca aatggatca 1620
 tatacataaaggaa acatatacgcc cttttttttt ggtttccctt gttttttttt gttttttttt 1680
 agggccaaagggt gggatggatc tttttttttt ggtttccctt gttttttttt gttttttttt 1740
 gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt 1800
 tagtcccttccat tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt 1860
 30 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1920
 gaataagaag aaggagaagg agaaggggaga agggaggggag aaggaggaggag gaggagaagg 1980
 agggatggatc gggatggatc gggatggatc gggatggatc gggatggatc gggatggatc 2040
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2100
 aqcccttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2160
 tagtactaga agtccatgtt agtccatgtt agtccatgtt agtccatgtt agtccatgtt 2220
 35 ctggaaaggaa agaagtccat tttttttttt tttttttttt tttttttttt tttttttttt 2280
 gactttaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2340
 caatgttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2400
 caaaaaaggca gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2460
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2520
 40 agatatttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2580
 agcaatttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2640
 agaaggaaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2700
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2760
 agtctatgtt acccaaacttccat tttttttttt tttttttttt tttttttttt tttttttttt 2820
 45 gaacagaata gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2880
 aggttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2940
 ctggatattccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3000
 aatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3060
 aacaccggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3120
 50 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3180
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3240
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3300
 ctcttataatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3360
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3420
 ctgtatcat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3480
 55 atggatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3540
 acccttggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3600
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3660
 caaaaaaggaaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3720
 60 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3780
 aaaaatgttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3840
 tcagttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3900
 aagacaaatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3960
 agaaatagatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4020
 65 taattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4080

<210> 5
 <211> 8616
 <212> DNA
 <213> Homo sapiens
 5
 <400> 5
 gtgaggaggt ggtggccgtc gaggggccag gccccagagc tgaatgcagt aggggctcag 60
 aaaaaggggc aggcagagcc ctggtcctcc tgcttcatac gtcacgtggg cacacgtggc 120
 10 ttttcgtca ggacgtcgg gggacacggt gatctctgcc tctgcctcc ctccgttcca 180
 gtttgcataa acttacgagg ttcaccttca cgttttgcgt gacacgggtt ttccaggcgc 240
 cgaggccaga gcagtgaaca gaggaggctg ggcgcggcag tggagccggg ttgccggcaa 300
 tggggagaag tgctggaa cacagacgt ctggcgaggg tgccgtcagg ttactataa 360
 tcttcgtca aatttcggg aatggggatgg gggatggggg cggaaacccc ctcttcgttgg 420
 15 ggggtggggggg taagggtttt gcaagggtcac gttgtcagcc aatatgcagg ttttgttta 480
 agattnaattt gtgtgttgc ggcagggtgc gttggctcac gccggtaatc ccagcaactt 540
 gggaaagctga ggcagggtggg tcacccgttca cggagggtt gggccggcgc 600
 ggtgaaacccc tattttttttt aatggggatgg gggatggggg tgccgtcagg 660
 taatccccc tattttttttt aatggggatgg gggatggggg tgccgtcagg 720
 tgcgtgagc tgagatgttcc cttttttttt cccatgttca cccatgttca cccatgttca 780
 20 ctttttttttt aaaaatgttcc cttttttttt cccatgttca cccatgttca cccatgttca 840
 actgttctcc agcacacatc ctggtcccat cttttttttt cccatgttca cccatgttca 900
 agaggacacgc agatggctcc acctgtcgttca gaaaggacacgc tttttttttt tttttttttt 960
 atgggtctgc tggggcccttgc cttttttttt cccatgttca cccatgttca cccatgttca 1020
 25 cttccgtcc acggccctttt tgggtcccttgc cttttttttt cccatgttca cccatgttca 1080
 gaaatcccgat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140
 cagacaaggg ggggtggccat tttttttttt tttttttttt tttttttttt tttttttttt 1200
 aaaaatgtca taacatgaga ttggcacttca taacacccgtt cttttttttt tttttttttt 1260
 tgcgttccatc cgggggtttt cccatgttca cccatgttca cccatgttca cccatgttca 1320
 30 gtccttccatcc atcgaacccgc agtgccttca cccatgttca cccatgttca cccatgttca 1380
 agtcagatataa ggcgttccatc accccatgttca cccatgttca cccatgttca cccatgttca 1440
 gagagtttgc gttttttttt cccatgttca cccatgttca cccatgttca cccatgttca 1500
 aggtcacaat ctggcccttgc cttatgttca ggggtggccat tttttttttt tttttttttt 1560
 cacgtgttccatc gttttttttt cccatgttca cccatgttca cccatgttca cccatgttca 1620
 35 ggcggccggcc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 1680
 cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 1740
 gagtgagggc tgggtcccttgc tgggtcccttgc tgggtcccttgc tgggtcccttgc tgggtcccttgc 1800
 ggggttcccttgc tgggtcccttgc tgggtcccttgc tgggtcccttgc tgggtcccttgc tgggtcccttgc 1860
 taggggttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 1920
 40 gtccttccatc gttttttttt cccatgttca cccatgttca cccatgttca cccatgttca 1980
 cacgtgttccatc gttttttttt cccatgttca cccatgttca cccatgttca cccatgttca 2040
 ggcgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2100
 cttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2160
 ggcgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2220
 45 cttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2280
 ggcgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2340
 ggcgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2400
 tttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2460
 ttgggttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2520
 50 cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2580
 gcaacgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2640
 ccagtcgtccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2700
 cttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2760
 ttgggttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2820
 55 gttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2880
 cttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 2940
 cactttcaag tttttttttt cccatgttca cccatgttca cccatgttca cccatgttca 3000
 ctgtgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 3060
 cttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 3120
 gttttttttt cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 3180
 60 ttatgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 3240
 ggagcccttgc tttttttttt cccatgttca cccatgttca cccatgttca cccatgttca 3300
 acatccgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 3360
 aagcttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 3420
 gggggatggg tttttttttt cccatgttca cccatgttca cccatgttca cccatgttca 3480
 65 gcatgttccatc cccatgttca cccatgttca cccatgttca cccatgttca cccatgttca 3540

gtccggatgg tgcaggctcg gggtaggtc gccaggccct tggtagactg gatgtgcgg 7500
 gtctgcatgg tgcaggctcg gggtaggtc gccaggccct tggtagactg gatgtgcgg 7560
 gtccggatgg tgcaggctcg gctgtggatc gccaggccct gctgtggatc ggatgtgcgg 7620
 tgctggatg tgcaggctcg gggtaggtc agccaaaggcc tccggtagc tggatgtgg 7680
 gtgtccggatgg tgcaggctcg cggggtaggtc tgcaggccct ctgggttagt ctggatgttc 7740
 ggtgtccggatgg tgcaggctcg cggggtaggtc gtcaccaggc cctgcggta gctggatgtc 7800
 cgggtgtctgg atggtagtgc tccggggtaggtc ggtcgccagg ccctgtgtc agctggatgt 7860
 gctgtatccg gatgtgcag tgcggggtaggtc aggtcgccag gccctgcgt gagetggatgt 7920
 tgctgtatccg gatgtgcag tgcggggtaggtc ggcggccct ggccctgcgg ttagctggatgt 7980
 atgggggtgc ggtatggtca ggttccacca gggccctgcgg ttagctggatgt 8040
 gtgcgggtgc cggatggtgc aggtctgggg tgaggctggcc agggccctgtc gtgagctgg 8100
 tggtagtgc tccggatgtgc cgggtccggg gtgaggctgc cagggccctgc ggtgagctgg 8160
 atgtgtgtatc tccggatgtgc cgggtccggg gtgaggctgc cagggccctgc cgggtgagctg 8220
 gatgtgcagt gtacggatgg tgcaggctcg gggtaggtc gccaggccct ggccgtgggt 8280
 gtatgtgtgt tgcaggatgg tgcaggatggcc ggggtggatgt cggccaggccc tgccggtagc 8340
 tggatgtgtgt tgcaggatgg tgcaggatggcc cggggtaggtc tgccggtagc ctccggtagc 8400
 tggatgtatccg tgcggggatgg tgcaggatgg tgcaggatggcc gggggtaggtc 8460
 tggtagtgc tgcggggatgg tgcaggatgg tgcaggatggcc gggggtaggtc 8520
 tggtagtgc tgcggggatgg tgcaggatgg tgcaggatggcc gggggtaggtc accaggccct 8580
 cgggtatctg gatgtggcat gtccttcctcg tttaag 8616

 <210> 6
 <211> 2089
 <212> DNA
 <213> Homo sapiens

 <400> 6
 gtactgtatc cccacgcggcc gcccggccctc ttcgaagtcc tggAACACCA gcccggccctc 60
 agcatgcggcc tgcggccact tgcgtgtgt tccctggctg tgcagctgtc ggctggggagc 120
 caggggccccc gtcacaggcc tggtccaagt ggattctgtg caaggctgtg actgccttgg 180
 gtcacgttc tcttaactgtt aaaaatcaggaa gtttggccaa agtgggttctt aggggttta 240
 aaggcagaagg gattttaaat agatggaaac actaccacta gctcccttgc ctccccctgg 300
 gatgtgggtc tgatcttc tctttttttttttttttt tgagatggatgt tctactgtc 360
 ttggccaggc tggatgtcgat tggataatc ttggctact gcaacatccca cctcttgggt 420
 ttaagcgatt caccaggcc tggatgtggaa ttacaggcac ctgcaccac 480
 gcctggctaa tttttgtact tttaggagag acgggggtttt accatgttgg ccaggctgg 540
 ctcgaactca tgacccctcagg tgatccaccc accttggccccc cccaaatgtc tgggttttaca 600
 ggctaaatccca cccggccccc cccggccatc tttttttttt catgtgttc tttatgtatc 660
 ttcaatctat tgatgttagg tcatgtggag ataaaaatccc accacttgc cgcactactg 720
 caggggacac ctgtcgaggg agcacctggg gataggagag ttccaccatg agtaacttc 780
 taggtggctg catttgaatg gctgtggat tttgtctca atgttggct gatgagatgt 840
 tgagattgtg acagattca gctggatttg catcgtgtgg gacggggagc gctgggttgc 900
 gagatgtccag cctgggtcgat cccaggccatc ggtttagtgc ttcctggc tccggccaggc 960
 tgactgtggaa gggcttttagt cagaagatca tggatgtttccccc agtccccccca cacaactgg 1020
 tccctgggggg gcttgc tgcacccatggc ccaaatcagg atgtctgc gacggggatgtgg 1080
 cagcagaccc tgcaggatgtt aacacagcc tggggctggg gacccggac gttgtgtgg 1140
 gccatcttc tgcacatgtggg ggagggatgtt ggccttccctt gttggaaacaa gttaaatcac 1200
 aatgcacccatc acttagactt tacacgtatt taatgtgtgt cggacccaaacaa tggtcatttgc 1260
 accgttattt tggaaaatgtttaatgggg tggatggggag gacggacac gatgtgggtt 1320
 ccccaagatg ctccctgtca ctactgggac tggatgtttctt cttggggggcc ctggggggcc 1380
 cttcttccttggggatgttcccttggggccctt tctactgttc tggggctggc gcctgcggc 1440
 agggcaccag ctccggagca cccggccccc cagtgccac gggatggccag gctgtcagcc 1500
 acagatgtccc aggtccaggatgttgc cagccccccgtt gccccccatgg gttgtttttgg 1560
 gggaaaaggc caaggccggatgg ggtgtcgatgg gactgggggg ctcatgtggatgttgc 1620
 tccctggggatgttcccttggggatgttcccttggggccctt cccggccctt ccatgtgttggatgtgggtt 1680
 cttttctactt gggggatgttcccttggggccctt ggccttgggttcccttggggccctt gtttcccttggggccctt 1740
 gggacaggca tccctgtgtgg agggggatgtt gttcacgtgg ccccaatgtc acgttggggac 1800
 caggctccctt ggtgtcgatgg gttggggacatgtt caccctgggg gttgaccggc gggactggggcc 1860
 tccccagggt tgcactatagg accagggtgtc cagggtggccctt gcaatgttggatgttggatgtgggtt 1920
 aggccggatgttcccttggggccctt gggggatgttcccttggggccctt gggggatgttcccttggggccctt 1980
 tgggtggccctt gggcccttcac tggatgttggatgttcccttggggccctt gggggatgttcccttggggccctt 2040
 tagttgttgc tctggatgttcccttggggccctt gggggatgttcccttggggccctt 2089

<210> 7
 <211> 687
 <212> DNA
 <213> Homo sapiens

<400> 7
 gtggctgtgc tttggtttaa cttccctttt aaacagaagt gcgtttgagc cccacatgg 60
 gtatcagttt agatgaaggg cccggaggag gggccacggg acacagccag ggcacatggca 120
 cggcgccaaac ccattttgtgc gcacagttag gttggccagg tgccgttgcc tccagaaaaag 180
 cagcgtgggg gtgttaggggg agtctctggg gcaaggacac gctctggaga ccacaagaag 240
 cagccggggcc agggccttggg tgcaacacgg cccgaggatcc tggtatccgtg tcctgtgtg 300
 gtgcgcagcc tccgtgcgtc tccgtttacg gggccgggg accaggccac gactggcagg 360
 agccccacgg gctctggagga tccctggaccc tgccccacgg ctctgtcacc ccacccctgt 420
 ggctggcggtt gtgcgggtt ccccggttccatc tgaggagagt gtgggttgag gtggacagag 480
 gtgtggcatg aggatccgt gtcaacacac catgcggcca ggaaccgggt tcaaacagg 540
 tcttggggaaat ctgggggggg ttcctggatcc cgggtctggg tggttgggaa cactggggag 600
 gggctgttcc tccccctgggtt ccctatgggtt ggggtggcac ttggccggat ccactttctt 660
 gactgtctcc catgctgtcc cccggccag 687

<210> 8
 <211> 494
 <212> DNA
 <213> Homo sapiens

<400> 8
 gtgggtggcg gggacccccc tgaggcagccc tgctggaccc tgggagtggc tgctgtatgg 60
 gcaacctatgtt tgggtggag gaggtactcc tgggtggcc gcaggaggatg caggtgaccc 120
 tggtaactgtt gaggacacac ctggcacacta ggggtggaggtt cttcagccctt tcctgcagca 180
 catggggccg actgtgcacc ctgactggcc gggctcttat tcccaaggag ggtttccactg 240
 gattcccaagt tccgtcagag aaggAACCGC aacggcttcag ccaccaggcc cccgtgcctt 300
 gcaaccccaagt cctgagccag ggggtctctg tcctggatcc cagaggggg acacagccgg 360
 ccctggccctt ggggtctggg tggttggggg tcagagagag agtgggggac accggccagg 420
 caggccctgtt gggcagaggt gatgtttggag tttctgtgtt gccactgtca gtctccctcg 480
 ctccactcaac acag 494

<210> 9
 <211> 865
 <212> DNA
 <213> Homo sapiens

<400> 9
 gtaagggttca cgtgtgtatag tctgtgtccag gatgtgtgtc tctggatata gaatgtgtct 60
 aaaaatcgat cgtgtgtgtg atgcgttctt gtgggtggagg tacttccatg attacacat 120
 ctgtgtatatgt cgtgtgtgtgc acgtgtgtgt cgtgtgtcat gtatctgtgg cgtgtcatatt 180
 tgggtgtgtgt gtgtgtgtgg cacgtgtgtg tccatgggtgt gtgtgtctgt ggtgtgtcatg 240
 tgggtgtgttc tggacacatgt gcatgttcat gctgtgtgtc gcatgtgtgt gatgtgtctta 300
 tttgtgtgtgt gtgtgtgtat gtgtccgtgtat cttatgtgtat tctatgtgtat ggtgtgtgt 360
 ggcaccccttgg ctttactactt tccatcttcca ggcatgttcc gcaatgtt cttccatgtt 420
 ctgggtgtgtgt gtttggggag ctccacatcc agggcttca tttcttagcat ggtgtccctt 480
 gtctgttcaac agggctggc cttggagact gtaagccagg tttgagagga gatgtggat 540
 gctgggttta ctttccctggaa cccctggcac ccccaaggacc ccagttgtgc ctatgcggcc 600
 tccatgtatgat ataggaaggc tgattcaggc ctcgtccccccc gggacacactt cttccacagg 660
 cgggggggggg ctttgggggtt cggcaggaggat gaaaggggcc ctgggtttgg gttttccccc 720
 agtgggtatgtt agcaacgtgg aggggtttaatc cttcaatgtc tgccaggccc ggggtgtcaga 780
 ggttggaaatg tttccctggaa gtttggatcc ttttggatggc acatgtggaa accccacaagg 840
 acctttttttt ctgactttttt gagct 865

<210> 10
 <211> 3782
 <212> DNA
 <213> Homo sapiens

<400> 10
 gtggggatgtt gttttcatgt gtggggatagg tggggatctg tggggatgggt tttttatgtgt 60
 ggggttaacac aqadttcaaq gcgagctttt ttcctgtatgtt ggggttgcag tggttccaaac 120

5 taggttagact gtcaaatactc agagcaaaat gaaaataaca aagttttaaa gggaaagaaa 2340
aaccccagct ctttggactt ccttaggct gaacttcate tcaagcagct tccttcac 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460
ggaggcgtg ggtgacacca gccaggaccc ctgaaaggaa gtggtgtgtt tcctgcctca 2520
gccccacgct cctgcccgtc ctgcacctgc tgaacegtc gatgtgtgtg ccaggtgccc 2580
acctggaaag gatgtgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640
gcacttgtgg caggcacaat tacagccct ccccaaagat gccccacgtcc ttcttcgtt 2700
acctgtgaat gtgtcacccg caaggcagag gctggtaag gctgcagggtg gaattcacggc 2760
tgccagtcag ccgatcttaa ggtcatctgg gattatctgg tgggcctgtatggccacaa 2820
gggtccctag aagtgagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
caactggccac tgctggctt gagatggagg agggggtccc cagccaaaggatg gggggcag 2940
ccgctccatg ctggaaaagc aagcaatctt cccggctctt gagggcacac ggcctgccc 3000
acgcctcgtat ttcaaggccag tgggacctgtt ttcagcttcc cggcctccag agctgtaa 3060
tgatgcgtt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
10 15 cagtagcaggaaatgaatac agggacagtt ctcagagtga ctctcagccc acccctggg 3179